

**PERAMALAN KUALITAS UDARA UNTUK PEMANTAUAN BERDASARKAN  
*WORLD AIR QUALITY INDEX* MENGGUNAKAN  
*SUPPORT VECTOR REGRESSION***

**SKRIPSI**

**Oleh:**

**MAULANA HAEKAL NOVAL AKBAR**

**NIM. 220605110149**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**PERAMALAN KUALITAS UDARA UNTUK PEMANTAUAN  
BERDASARKAN *WORLD AIR QUALITY INDEX*  
MENGUNAKAN *SUPPORT VECTOR REGRESSION***

**SKRIPSI**

Diajukan kepada:

Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

**Oleh:**

**MAULANA HAEKAL NOVAL AKBAR**  
**NIM. 220605110149**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**HALAMAN PERSETUJUAN**

**PERAMALAN KUALITAS UDARA UNTUK PEMANTAUAN  
BERDASARKAN *WORLD AIR QUALITY INDEX*  
MENGUNAKAN *SUPPORT VECTOR REGRESSION***

**SKRIPSI**

**Oleh:**  
**MAULANA HAEKAL NOVAL AKBAR**  
**NIM. 220605110149**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 10 April 2026

Pembimbing I,



Dr. Agung Teguh Wibowo Almais, M.T.  
NIP. 19860301 202321 1 016

Pembimbing II,



Dr. M. Ainul Yaqin, M.Kom.  
NIP. 19761013 200604 1 004

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom  
NIP. 19841010 201903 1 012

## HALAMAN PENGESAHAN

### PERAMALAN KUALITAS UDARA UNTUK PEMANTAUAN BERDASARKAN *WORLD AIR QUALITY INDEX* MENGUNAKAN *SUPPORT VECTOR REGRESSION*

#### SKRIPSI

Oleh:  
**MAULANA HAEKAL NOVAL AKBAR**  
NIM. 220605110149

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal: 11 Maret 2026

#### Susunan Dewan Penguji

Ketua Penguji : Dr. M. Amin Hariyadi, M.T.  
NIP. 19670118 200501 1 001



Anggota Penguji I : Fajar Rohman Hariri, M.Kom.  
NIP. 19890515 201801 1 001



Anggota Penguji II : Dr. Agung Teguh Wibowo Almais, M.T.  
NIP. 19860301 202321 1 016



Anggota Penguji III : Dr. M. Ainul Yaqin, M.Kom.  
NIP. 19761013 200604 1 004



Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom  
NIP. 19841010 201903 1 012

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Maulana Haekal Noval Akbar  
NIM : 220605110149  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Peramalan Kualitas Udara Untuk Pemantauan  
Berdasarkan *World Air Quality Index*  
Menggunakan *Support Vector Regression*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 22 April 2026  
Yang membuat pernyataan,



Maulana Haekal Noval Akbar  
NIM.220605110149

## **MOTTO**

*“Setiap masalah adalah peluang untuk menemukan metode yang lebih baik”*

## HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, karya ini penulis persembahkan kepada:

Ayah, Ibu, dan kakak

atas doa, dukungan, dan kasih sayang tanpa henti yang selalu menjadi sumber  
kekuatan dan motivasi.

Dosen pembimbing dan para pengajar

atas bimbingan, arahan, serta ilmu yang sangat berarti bagi penulis.

Teman-teman Infinity Teknik Informatika 2022

yang telah menjadi bagian dari perjalanan ini melalui kebersamaan dan semangat  
yang tidak pernah padam.

+

Dan akhirnya, untuk diri sendiri

yang terus berusaha, melangkah, dan bertahan hingga mencapai titik ini.

## KATA PENGANTAR

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Alhamdulillahirabbil'alamin, segala puji syukur penulis panjatkan ke hadirat Allah SWT. atas limpahan rahmat, nikmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tugas skripsi yang berjudul “PERAMALAN KUALITAS UDARA UNTUK PEMANTAUAN BERDASARKAN *WORLD AIR QUALITY INDEX* MENGGUNAKAN *SUPPORT VECTOR REGRESSION*” dengan baik dan lancar. Shalawat serta salam tak lupa penulis sampaikan kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam, sosok teladan mulia yang telah menuntun kita dari zaman kegelapan menuju zaman kebenaran, Islam, dan ilmu pengetahuan yang kita rasakan manfaatnya hingga hari ini.

Tugas skripsi ini disusun dalam rangka memenuhi salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang. Dalam proses penyusunan tugas ini, penulis mendapatkan bantuan, dukungan, serta doa dari banyak pihak, baik secara langsung maupun tidak langsung. Untuk itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
2. Dr. H. Agus Mulyono, S.Pd., M.Kes, selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi S1 Teknik Informatika UIN Maulana Malik Ibrahim Malang sekaligus Dosen Wali.



4. Dr. Agung Teguh Wibowo Almais, M.T., selaku Dosen Pembimbing I, yang dengan kesabaran dan dedikasinya membimbing penulis selama penyusunan skripsi.
5. Dr. M. Ainul Yaqin, M.Kom., selaku Dosen Pembimbing II, yang senantiasa memberikan arahan dan masukan yang sangat berharga dalam pengerjaan skripsi ini.
6. Dr. M. Amin Hariyadi, M.T., dan Fajar Rohman Hariri, M.Kom., selaku Dosen Penguji yang telah memberikan kritik, saran, dan bimbingan untuk menyempurnakan skripsi ini.
7. Nia Faricha, S.Si., admin Program Studi Teknik Informatika, yang membantu dalam urusan administrasi dan selalu mengingatkan tentang kelengkapan berkas.
8. Bapak Nasihatul Munir dan Ibu Sholihatul Mir'ati, orang tua tercinta, serta Muhammad Haidar Zamzam Rijaldi, kakak saya, yang senantiasa menjadi sumber kekuatan melalui doa, kasih sayang, dan dukungan tanpa henti.
9. Seluruh keluarga, teman, sahabat, dan kerabat penulis yang tidak dapat penulis sebutkan satu per satu, yang turut memberikan bantuan, semangat, dukungan, dan doa untuk penulis.

Dan terakhir tak lupa terimakasih kepada diri sendiri, yang dengan tekad kuat mampu melewati setiap tantangan, rintangan, dan kesulitan selama perjalanan ini, membuktikan bahwa kerja keras dan usaha tidak pernah sia-sia. Penulis menyadari bahwa penelitian dalam tugas skripsi ini masih jauh dari kata sempurna dan memiliki berbagai keterbatasan. Oleh karena itu, dengan penuh kerendahan

hati, penulis membuka diri untuk menerima kritik dan saran yang membangun dari para pembaca guna menjadi bahan evaluasi dan pengembangan di masa mendatang. Penelitian ini juga memiliki potensi untuk dilanjutkan dan dikembangkan lebih jauh dalam penelitian berikutnya, sehingga dapat melengkapi kekurangan yang ada. Penulis berharap, karya ini tidak hanya memberikan manfaat bagi pembaca, tetapi juga dapat berkontribusi positif bagi masyarakat secara luas.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh*

Malang, 22 April 2026

Penulis

## DAFTAR ISI

<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN .....</b>	<b>v</b>
<b>MOTTO .....</b>	<b>vi</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>ABSTRAK .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>البحث مستخلص.....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian .....	7
1.5 Manfaat Penelitian .....	8
<b>BAB II STUDI PUSTAKA .....</b>	<b>9</b>
2.1 Penelitian Terkait dengan SVR dan Kualitas Udara .....	9
2.2 Pengantar Kualitas Udara.....	14
2.3 Indeks Kualitas Udara .....	15
2.4 WAQI API sebagai Sumber Data Kualitas Udara .....	18
2.5 <i>Support Vector Regression</i> (SVR) dalam Peramalan Kualitas Udara .....	19
2.6 Transformasi <i>Yeo-Johnson</i> pada Variabel Target .....	22
2.7 Validasi dan Evaluasi Model Peramalan.....	24
<b>BAB III DESAIN DAN IMPLEMENTASI .....</b>	<b>27</b>
3.1 <i>Data Preparation</i> .....	27
3.1.1 Sumber dan Variabel.....	27
3.1.2 Akuisisi Data.....	28
3.1.3 Penyimpanan Data .....	29
3.1.4 Praproses .....	31
3.1.5 Rekayasa Fitur.....	32
3.2 Alur dan Desain Sistem.....	33
3.2.1 Arsitektur <i>End-to-End</i> .....	33
3.2.2 Proses di Flask .....	37
3.3 Skenario Pengujian.....	40
3.3.1 Eksperimen Terbatas .....	41
3.3.2 Evaluasi Model.....	42

3.4 Perhitungan Manual .....	43
3.4.1 Data Awal.....	44
3.4.2 Rekayasa Fitur.....	45
3.4.3 Normalisasi Data dan Seleksi Fitur.....	48
3.4.4 Pemodelan dan Peramalan .....	51
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>53</b>
4.1 Gambaran Umum Penelitian .....	53
4.2 Karakteristik Data Penelitian .....	54
4.2.1 Sumber dan Rentang Data.....	55
4.2.2 Variabel yang Digunakan.....	57
4.2.3 Resolusi Temporal Data .....	58
4.3 Pra-pemrosesan dan Rekayasa Fitur .....	60
4.3.1 Pra-pemrosesan Data.....	60
4.3.2 Rekayasa Fitur.....	62
4.3.3 Seleksi Fitur .....	63
4.4 Implementasi Model <i>Support Vector Regression</i> .....	69
4.4.1 Konfigurasi Model .....	70
4.4.2 Skema Validasi.....	71
4.5 Implementasi Sistem Informasi Berbasis Web .....	72
4.6 Hasil Evaluasi Model .....	75
4.6.1 Metrik Evaluasi .....	76
4.6.2 Hasil Evaluasi Model SVR Per Wilayah dan Perbandingan dengan <i>Baseline</i> .....	78
4.6.3 Hasil Eksperimen Terbatas .....	83
4.7 Pembahasan Hasil .....	107
4.8 Integrasi Islam .....	111
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>114</b>
5.1 Kesimpulan .....	114
5.2 Saran .....	115
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 3.1 Diagram Hubungan Entitas Basis Data.....	30
Gambar 3.2 Arsitektur <i>end-to-end</i> sistem peramalan berbasis SVR.....	35
Gambar 3.3 Alur proses peramalan PM2.5 di Flask .....	38
Gambar 4.1 Halaman Utama Website.....	73
Gambar 4.2 Halaman Detail Wilayah Website .....	74
Gambar 4.3 Perbandingan Distribusi R <sup>2</sup> : SVR vs <i>Baseline</i> .....	80
Gambar 4.4 Perbandingan Distribusi MAE: SVR vs <i>Baseline</i> .....	81
Gambar 4.5 Perbandingan Distribusi RMSE: SVR vs <i>Baseline</i> .....	82
Gambar 4.6 Perbandingan Distribusi Nilai R <sup>2</sup> Model SVR pada Konfigurasi Panjang <i>Lag</i> dan <i>Rolling Window</i> .....	86
Gambar 4.7 Perbandingan Distribusi Nilai MAE Model SVR pada Konfigurasi Panjang <i>Lag</i> dan <i>Rolling Window</i> .....	87
Gambar 4.8 Perbandingan Distribusi Nilai RMSE Model SVR pada Konfigurasi Panjang <i>Lag</i> dan <i>Rolling Window</i> .....	88
Gambar 4.9 Perbandingan Distribusi Nilai R <sup>2</sup> Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial.....	92
Gambar 4.10 Perbandingan Distribusi Nilai MAE Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial .....	93
Gambar 4.11 Perbandingan Distribusi Nilai RMSE Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial .....	94
Gambar 4.12 Perbandingan Distribusi Nilai R <sup>2</sup> Model SVR Tanpa Seleksi Fitur, <i>SelectKBest</i> k=5, dan k=8.....	98
Gambar 4.13 Perbandingan Distribusi Nilai MAE Model SVR Tanpa Seleksi Fitur, <i>SelectKBest</i> k=5, dan k=8.....	99
Gambar 4.14 Perbandingan Distribusi Nilai RMSE Model SVR Tanpa Seleksi Fitur, <i>SelectKBest</i> k=5, dan k=8.....	100
Gambar 4.15 Perbandingan Distribusi Nilai R <sup>2</sup> Model SVR Tanpa Transformasi Target dan dengan Transformasi <i>Yeo-Johnson</i> .....	104
Gambar 4.16 Perbandingan Distribusi Nilai MAE Model SVR Tanpa Transformasi Target dan dengan Transformasi <i>Yeo-Johnson</i> .....	105
Gambar 4.17 Perbandingan Distribusi Nilai RMSE Model SVR Tanpa Transformasi Target dan dengan Transformasi <i>Yeo-Johnson</i> .....	106

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait .....	9
Tabel 2.2 <i>Breakpoints</i> US EPA PM2.5 .....	16
Tabel 2.3 <i>Breakpoints</i> ISPU KLHK PM2.5 .....	16
Tabel 3.1 Lokasi Stasiun WAQI yang Digunakan dalam Penelitian .....	28
Tabel 3.2 Contoh Data <i>Dummy</i> .....	45
Tabel 3.3 Hasil Seleksi Fitur terhadap Target (Data <i>Dummy</i> ) .....	49
Tabel 4.1 Ringkasan Data Penelitian per Wilayah .....	56
Tabel 4.2 Variabel Penelitian yang Digunakan dalam Model Peramalan .....	57
Tabel 4.3 Pola Fitur Terpilih pada Beberapa Wilayah .....	64
Tabel 4.4 Hasil Evaluasi Model SVR Per Wilayah dan Perbandingan dengan <i>Baseline</i> .....	78
Tabel 4.5 Ringkasan Konfigurasi Terbaik pada Eksperimen 1: Variasi <i>Lag</i> dan <i>Rolling Window</i> .....	85
Tabel 4.6 Ringkasan Konfigurasi Terbaik pada Eksperimen 2: Perbandingan SVR dengan dan tanpa Pembobotan Waktu .....	91
Tabel 4.7 Ringkasan Konfigurasi Terbaik pada Eksperimen 3: Perbandingan Tanpa Seleksi Fitur dan <i>SelectKBest</i> .....	97
Tabel 4.8 Ringkasan Konfigurasi Terbaik pada Eksperimen 4: Pengaruh Transformasi Target <i>Yeo-Johnson</i> .....	102

## ABSTRAK

Akbar, Maulana Haekal Noval. 2026. **Peramalan Kualitas Udara Untuk Pemantauan Berdasarkan *World Air Quality Index* Menggunakan *Support Vector Regression***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Agung Teguh Wibowo Almais, M.T. (II) Dr. M. Ainul Yaqin, M.Kom.

**Kata kunci:** PM2.5, Peramalan, *Support Vector Regression*, *World Air Quality Index*, Kualitas Udara, Sistem Informasi

Kualitas udara merupakan faktor penting yang memengaruhi kesehatan masyarakat dan kondisi lingkungan. Konsentrasi partikulat halus PM2.5 menjadi salah satu indikator utama pencemaran udara karena dapat meningkatkan risiko gangguan pernapasan dan penyakit kardiovaskular. Oleh karena itu, diperlukan metode yang mampu meramalkan kualitas udara untuk memberikan informasi awal mengenai kondisi udara di masa mendatang. Penelitian ini bertujuan mengembangkan model peramalan konsentrasi PM2.5 menggunakan metode *Support Vector Regression* (SVR) serta mengimplementasikannya dalam sistem informasi pemantauan kualitas udara berbasis web. Data penelitian diperoleh dari *World Air Quality Index* (WAQI) API yang mencakup lima belas wilayah di Indonesia dengan periode pengamatan Juni 2025 hingga Januari 2026. Variabel yang digunakan meliputi PM2.5 sebagai variabel dependen serta variabel meteorologi berupa temperatur, kelembapan, tekanan udara, kecepatan angin, dan titik embun sebagai variabel independen. Proses pemodelan meliputi pra-pemrosesan data, rekayasa fitur, seleksi fitur menggunakan *SelectKBest*, serta pelatihan model SVR dengan kernel *Radial Basis Function*. Evaluasi model menggunakan metrik *Mean Absolute Error*, *Root Mean Squared Error*, dan koefisien determinasi. Hasil penelitian menunjukkan bahwa model SVR menghasilkan performa peramalan yang lebih baik dibandingkan model *baseline persistence* dengan nilai  $R^2$  antara 0.089 hingga 0.512. Sistem informasi yang dikembangkan mampu menampilkan data kualitas udara dan hasil peramalan secara otomatis.

## ABSTRACT

Akbar, Maulana Haekal Noval. 2026. Air Quality Forecasting for Monitoring Based on the World Air Quality Index Using Support Vector Regression. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Dr. Agung Teguh Wibowo Almais, M.T. (II) Dr. M. Ainul Yaqin, M.Kom.

**Key words:** PM2.5, Forecasting, Support Vector Regression, World Air Quality Index, Air Quality, Information System

Air quality is a critical factor affecting public health and environmental conditions. Concentrations of fine particulate matter (PM2.5) are a key indicator of air pollution because they can increase the risk of respiratory disorders and cardiovascular diseases. Therefore, a method capable of forecasting air quality is needed to provide early information on future air conditions. This study aims to develop a PM2.5 concentration forecasting model using the Support Vector Regression (SVR) method and to implement it in a web-based air quality monitoring information system. Research data were obtained from the World Air Quality Index (WAQI) API, covering fifteen regions in Indonesia with an observation period from June 2025 to January 2026. The variables used include PM2.5 as the dependent variable and meteorological variables such as temperature, humidity, air pressure, wind speed, and dew point as independent variables. The modeling process includes data preprocessing, feature engineering, feature selection using *SelectKBest*, and training the SVR model with a *Radial Basis Function* kernel. Model evaluation uses the metrics *Mean Absolute Error*, *Root Mean Squared Error*, and the coefficient of determination. The results show that the SVR model produces better forecasting performance compared to the *baseline persistence* model, with  $R^2$  values ranging from 0.089 to 0.512. The developed information system is capable of automatically displaying air quality data and forecasting results.



## البحث مستخلص

أكبر، مولانا هيكل نوفال. 2026 توقعات جودة الهواء للرصد استنادًا إلى مؤشر جودة الهواء العالمي باستخدام انحدار المتجه الداعم. أطروحة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرفون : (الأول) الدكتور أغونغ تيغوه ويبوو ألبايس، ماجستير في التقنية (الثاني) الدكتور م. عين اليقين، ماجستير في علوم الحاسوب.

الكلمات المفتاحية: PM2.5، التنبؤ، انحدار المتجهات الداعمة، مؤشر جودة الهواء العالمي، جودة الهواء، نظام المعلومات

تعد جودة الهواء عاملاً مهمًا يؤثر على صحة المجتمع وحالة البيئة. وتعد تركيزات الجسيمات الدقيقة PM2.5 أحد المؤشرات الرئيسية لتلوث الهواء، حيث إنها تزيد من مخاطر الإصابة بمشاكل الجهاز التنفسي وأمراض القلب والأوعية الدموية. ولذلك، هناك حاجة إلى طريقة قادرة على التنبؤ بجودة الهواء لتوفير معلومات مسبقة حول حالة الهواء في المستقبل. تهدف هذه الدراسة إلى تطوير نموذج للتنبؤ بتركيز PM2.5 باستخدام طريقة *Support Vector Regression (SVR)* وتطبيقه في نظام معلومات مراقبة جودة الهواء على شبكة الإنترنت. تم الحصول على بيانات الدراسة من واجهة برمجة تطبيقات *World Air Quality Index (WAQI)* التي تغطي خمسة عشر منطقة في إندونيسيا، مع فترة رصد تمتد من يونيو 2025 إلى يناير 2026. تشمل المتغيرات المستخدمة PM2.5 كمتغير تابع، بالإضافة إلى المتغيرات الجوية مثل درجة الحرارة والرطوبة وضغط الهواء وسرعة الرياح ونقطة الندى كمتغيرات مستقلة. تشمل عملية النمذجة المعالجة المسبقة للبيانات، وهندسة الميزات، واختيار الميزات باستخدام *SelectKBest*، بالإضافة إلى تدريب نموذج *SVR* باستخدام نواة *Radial Basis Function*. تم تقييم النموذج باستخدام مقاييس *Mean Absolute Error* و *Root Mean Squared Error* ومعامل التحديد. أظهرت نتائج البحث أن نموذج *SVR* حقق أداءً تنبؤيًا أفضل مقارنةً بنموذج *baseline persistence* بقيمة  $R^2$  تتراوح بين 0.089 و 0.512. تمكن نظام المعلومات الذي تم تطويره من عرض بيانات جودة الهواء ونتائج التنبؤ تلقائيًا.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kualitas udara merupakan permasalahan global yang mendapat perhatian luas karena berdampak langsung terhadap kesehatan manusia, aktivitas ekonomi, dan kenyamanan hidup masyarakat (Umah & Gusmira, 2024). Aktivitas transportasi, industri, serta kegiatan rumah tangga meningkatkan konsentrasi *Particulate Matter* berdiameter  $\leq 2,5$  mikrometer (PM2.5) di atmosfer. Partikel ini berbahaya karena dapat masuk ke dalam sistem pernapasan manusia dan memicu berbagai penyakit seperti gangguan pernapasan, penyakit kardiovaskular, hingga kematian dini apabila terpapar dalam jangka Panjang (Haya et al., 2025). World Health Organization (WHO) menyatakan bahwa polusi udara menjadi salah satu penyebab utama jutaan kematian prematur setiap tahun di seluruh dunia (Haya et al., 2025). Selain dampak kesehatan, kualitas udara yang buruk juga mempengaruhi kesejahteraan masyarakat melalui peningkatan biaya kesehatan, penurunan produktivitas kerja, serta gangguan aktivitas sosial dan ekonomi. Penelitian terbaru menunjukkan bahwa indeks kesehatan dan kesejahteraan masyarakat memiliki korelasi negatif dengan penurunan kualitas udara (Haya et al., 2025). Kondisi tersebut menunjukkan bahwa upaya pengendalian kualitas udara menjadi bagian penting dalam mendukung pembangunan berkelanjutan.

Indonesia termasuk negara yang rentan terhadap degradasi kualitas udara akibat karakteristik geografis dan tingginya aktivitas ekonomi (Yunita *et al.*, 2023).

Peningkatan jumlah kendaraan bermotor di kota-kota besar seperti Jakarta, Surabaya, dan Medan berkontribusi terhadap meningkatnya emisi gas buang dan pencemaran udara perkotaan (Umah & Gusmira, 2024). Selain itu, kebakaran hutan dan lahan gambut di wilayah Sumatera dan Kalimantan hampir setiap tahun menimbulkan kabut asap yang berdampak luas terhadap kesehatan masyarakat serta mengganggu aktivitas transportasi udara dan jarak pandang (R. A. Rahman *et al.*, 2024). Fenomena tersebut tidak hanya menimbulkan dampak domestik, tetapi juga mempengaruhi negara-negara tetangga melalui penyebaran asap lintas batas. Oleh karena itu, diperlukan strategi mitigasi yang efektif untuk mengurangi dampak polusi udara dan meningkatkan kualitas udara secara berkelanjutan.

Dampak polusi udara juga memberikan tekanan signifikan terhadap perekonomian. Peningkatan kasus penyakit akibat polusi udara menyebabkan meningkatnya biaya kesehatan serta menurunkan produktivitas masyarakat (Haya *et al.*, 2025). Ketika kualitas udara memburuk, aktivitas luar ruang cenderung menurun sehingga berdampak pada sektor pariwisata, pendidikan, dan jasa. Selain itu, kabut asap juga dapat mengganggu lalu lintas udara yang berimbas pada rantai pasokan nasional (Zhang *et al.*, 2022). Penelitian terbaru menunjukkan bahwa polusi udara di kawasan Asia Tenggara menyebabkan kerugian ekonomi hingga miliaran dolar setiap tahun (Anastasya & Suwandana, 2022). Kondisi ini menunjukkan bahwa sistem pemantauan dan peramalan kualitas udara sangat penting untuk mendukung upaya mitigasi risiko kesehatan sekaligus mengurangi kerugian ekonomi akibat polusi udara. Selain dari perspektif ilmiah dan sosial ekonomi, permasalahan polusi PM2.5 juga dapat dipandang sebagai konsekuensi

dari aktivitas manusia dalam memanfaatkan sumber daya alam tanpa pengelolaan yang bertanggung jawab. Fenomena ini sejalan dengan firman Allah SWT dalam Surah Ar-Rum Ayat 41:

ظَهَرَ الْفَسَادُ فِي الْبَرِّ وَالْبَحْرِ بِمَا كَسَبَتْ أَيْدِي النَّاسِ لِيُذِيقَهُمْ بَعْضَ الَّذِي عَمِلُوا لَعَلَّهُمْ يَرْجِعُونَ

*“Telah nampak kerusakan di darat dan di laut disebabkan karena perbuatan tangan manusia, supaya Allah merasakan kepada mereka sebagian dari (akibat) perbuatan mereka, agar mereka kembali (ke jalan yang benar)” (QS. Ar-Rum:41).*

Ibnu Katsir menjelaskan bahwa kerusakan yang terjadi di darat dan laut merupakan akibat dari perbuatan manusia, yang dapat terlihat dari berkurangnya keberkahan alam, menurunnya hasil tanaman, serta berbagai kesulitan yang dihadapi manusia (Shaykh & Ghoffar, 2006). Oleh karena itu, penelitian mengenai peramalan kualitas udara dapat dipandang sebagai salah satu upaya ilmiah untuk memahami dampak aktivitas manusia terhadap lingkungan sekaligus mendorong pengelolaan lingkungan yang lebih bertanggung jawab.

Penelitian ini memanfaatkan data terbuka yang diperoleh melalui *Application Programming Interface (API) World Air Quality Index (WAQI)*. WAQI merupakan platform yang mengagregasi data dari berbagai stasiun pemantauan kualitas udara yang terafiliasi dengan lembaga resmi seperti Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) (The World Air Quality Index, n.d.). WAQI API menyediakan data berkala yang dapat dimanfaatkan dalam pengembangan sistem pemantauan dan peramalan kualitas udara (Joshi *et al.*, 2025). Dalam penelitian ini, data historis WAQI digunakan untuk membangun sistem peramalan kualitas udara jangka pendek dengan horizon H+1, yaitu satu hari kalender berikutnya. Karakteristik data tersebut menunjukkan bahwa interval

pengamatan tidak selalu seragam karena dalam satu hari dapat terdapat lebih dari satu pengamatan atau hanya satu pengamatan, tergantung pada ketersediaan data dari stasiun WAQI. Selain itu, data WAQI juga memiliki keterbatasan seperti keterlambatan pembaruan data akibat gangguan sensor atau transmisi serta tidak tersedianya beberapa variabel eksternal seperti kebakaran hutan atau aktivitas transportasi (Az-Zahran *et al.*, 2024).

Sebagian besar penelitian sebelumnya mengenai peramalan kualitas udara masih berfokus pada satu wilayah atau satu stasiun pemantauan tertentu. Pendekatan tersebut menyebabkan hasil penelitian hanya merepresentasikan kondisi lokal dan belum menggambarkan performa model pada karakteristik lingkungan yang berbeda. Oleh karena itu, penelitian ini memanfaatkan data WAQI API untuk melakukan peramalan konsentrasi PM<sub>2.5</sub> pada 15 wilayah di Indonesia secara terpisah sehingga memungkinkan evaluasi performa model pada berbagai kondisi lingkungan.

Metode yang digunakan dalam penelitian ini adalah *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF). Pemilihan metode tersebut didasarkan pada kemampuan SVR dalam memodelkan hubungan non-linear antara variabel input dan output. Konsentrasi PM<sub>2.5</sub> dipengaruhi oleh berbagai variabel meteorologi seperti suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun yang memiliki hubungan kompleks dan tidak linear. Kernel RBF memungkinkan model SVR menangkap pola hubungan tersebut secara lebih efektif dibandingkan model linear (Chen *et al.*, 2020). Penelitian sebelumnya juga menunjukkan bahwa SVR dengan kernel RBF mampu memberikan performa

peramalan yang baik dengan nilai *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) yang rendah serta stabil terhadap data (Maltare & Vahora, 2023).

Penelitian ini tidak hanya berfokus pada pengembangan model peramalan, tetapi juga mengembangkan *pipeline* sistem secara *end-to-end* yang mencakup proses pengambilan data dari WAQI API, praproses data, pelatihan model, serta evaluasi performa model (Meisenbacher *et al.*, 2022). *Pipeline* tersebut kemudian diintegrasikan dengan sistem informasi berbasis web sehingga hasil peramalan dapat ditampilkan secara langsung kepada pengguna sebagai informasi kondisi kualitas udara dan prakiraan kualitas udara pada hari berikutnya.

Pengembangan sistem peramalan kualitas udara berbasis *machine learning* juga memiliki relevansi terhadap tujuan pembangunan berkelanjutan atau *Sustainable Development Goals* (SDGs). Sistem ini mendukung SDG 3, khususnya target 3.9, yaitu mengurangi jumlah penyakit dan kematian akibat polusi udara (United Nations Statistics Division, 2023a). Selain itu, sistem ini juga berkaitan dengan SDG 11, khususnya target 11.6, yang bertujuan mengurangi dampak lingkungan perkotaan termasuk peningkatan kualitas udara (United Nations Statistics Division, 2023b). Studi menunjukkan bahwa polusi udara di Jakarta telah menyebabkan ribuan kasus penyakit pada anak-anak, lebih dari sepuluh ribu kematian, serta kerugian ekonomi lebih dari USD 2,9 miliar per tahun atau sekitar Rp42,9 triliun (Syuhada *et al.*, 2023). Data tersebut menunjukkan bahwa pengendalian polusi udara merupakan upaya penting dalam mendukung pembangunan berkelanjutan serta memperkuat relevansi penelitian ini.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat diidentifikasi beberapa permasalahan dalam penelitian ini sebagai berikut:

1. Pemanfaatan data kualitas udara yang diperoleh melalui *Application Programming Interface (API) World Air Quality Index (WAQI)* masih belum optimal dalam membangun dataset penelitian yang terstruktur dan siap digunakan untuk proses peramalan kualitas udara.
2. Penerapan metode *Support Vector Regression (SVR)* dengan kernel *Radial Basis Function (RBF)* dalam memodelkan peramalan konsentrasi PM<sub>2.5</sub> masih memerlukan pengembangan agar menghasilkan peramalan yang akurat.
3. Perancangan *pipeline* sistem yang mencakup proses pengambilan data dari WAQI API, praproses data, pelatihan model, hingga penyajian hasil peramalan kualitas udara pada 15 wilayah di Indonesia masih belum terintegrasi secara optimal.

## 1.3 Batasan Masalah

Penulis menetapkan batasan masalah berikut agar penelitian ini lebih fokus dan terarah:

1. Penelitian ini memanfaatkan data kualitas udara historis yang diperoleh melalui WAQI, yang berfungsi sebagai agregator data dari berbagai stasiun pemantauan. Sebagian stasiun tersebut diketahui memiliki atribusi kepada lembaga resmi di Indonesia seperti BMKG, berdasarkan informasi metadata yang tersedia.

2. Penelitian ini terbatas pada penggunaan data historis WAQI dengan frekuensi pembaruan yang tidak selalu konsisten, sehingga dapat memengaruhi akurasi hasil peramalan. Selain itu, metode yang digunakan hanya *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF), tanpa melakukan perbandingan dengan algoritma *machine learning* lainnya.
3. Peramalan hanya dilakukan pada lima belas lokasi stasiun pemantau yang terpilih, dengan menggunakan variabel meteorologi lokal yang tersedia seperti suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun. Model ini memiliki keterbatasan dalam menangkap pengaruh faktor eksternal berskala regional, seperti asap kebakaran hutan atau emisi lokal spesifik yang tidak terdeteksi oleh stasiun pemantauan.

#### **1.4 Tujuan Penelitian**

1. Mengolah data kualitas udara yang diperoleh dari WAQI API sehingga dapat digunakan sebagai dataset untuk analisis dan peramalan kualitas udara.
2. Membangun model peramalan konsentrasi PM<sub>2.5</sub> menggunakan metode *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF).
3. Mengembangkan *pipeline* sistem yang mampu mengintegrasikan proses pengambilan data, praproses data, pelatihan model, serta penyajian hasil peramalan kualitas udara pada 15 wilayah di Indonesia.



## 1.5 Manfaat Penelitian

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode peramalan kualitas udara menggunakan pendekatan *machine learning*, khususnya *Support Vector Regression (SVR)*, serta menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan pemanfaatan data kualitas udara berbasis API.
2. Penelitian ini diharapkan dapat menghasilkan sistem peramalan kualitas udara yang mampu memberikan informasi mengenai kondisi udara dan prakiraan kualitas udara pada hari berikutnya (H+1) sehingga dapat membantu masyarakat maupun pihak terkait dalam meningkatkan kewaspadaan terhadap risiko polusi udara.
3. Penelitian ini diharapkan dapat mendukung upaya pemantauan kualitas udara serta meningkatkan kesadaran masyarakat terhadap pentingnya menjaga kualitas lingkungan hidup sebagai bagian dari upaya mewujudkan pembangunan berkelanjutan.

## BAB II

### STUDI PUSTAKA

#### 2.1 Penelitian Terkait dengan SVR dan Kualitas Udara

Sejumlah penelitian telah dilakukan untuk memanfaatkan pendekatan *machine learning*, khususnya *Support Vector Regression* (SVR), dalam meramalkan kualitas udara. Berbagai studi menunjukkan bahwa SVR memiliki keunggulan dalam menangani hubungan *non-linear* antara polutan udara dan variabel meteorologi, serta mampu memodelkan fluktuasi deret waktu yang dinamis. Metode ini juga sering dibandingkan dengan pendekatan lain berbasis optimasi untuk meningkatkan akurasi peramalan. Tabel 2.1 berikut merangkum beberapa penelitian terkait penggunaan SVR dalam peramalan kualitas udara.

Tabel 2.1 Penelitian Terkait

No.	Referensi	Topik	Metode	Subjek	Hasil
1.	(Shafii <i>et al.</i> , 2020)	Peramalan <i>Air Pollution Index</i> (API) PM2.5 menggunakan SVR	SVR dengan kernel RBF, dibandingkan dengan tiga kernel lainnya.	Data dari CAQM di Tanah Merah dan Kota Bharu, Malaysia.	Kernel RBF dengan cost dan gamma 100 menghasilkan MAE dan RMSE rendah, menunjukkan performa terbaik
2.	(Houssein <i>et al.</i> , 2025)	Peramalan kualitas udara PM2.5 menggunakan	Penggabungan HHO untuk optimasi parameter SVR dalam	Dataset PM2.5 dari model Downscaler Badan Perlindungan	HHO-SVR unggul dibandingkan GWO

		an model hybrid <i>Harris Hawks Optimization</i> (HHO) dan SVR	meramalkan konsentrasi PM2.5	Lingkungan AS (EPA)	dan WOA dalam akurasi dan waktu komputasi
3.	(Abuouelezz <i>et al.</i> , 2025)	Peramalan PM2.5 dan PM <sub>10</sub> di Abu Dhabi menggunakan <i>machine learning</i>	<i>Decision Tree</i> (DT), <i>Random Forest</i> (RF), <i>Support Vector Regression</i> (SVR), <i>Convolutional Neural Network</i> (CNN), <i>Long Short-Term Memory</i> (LSTM), dan <i>Facebook Prophet</i>	Data lima tahun dari enam stasiun di Abu Dhabi.	SVR terbaik untuk PM2.5 jangka pendek, CNN untuk PM <sub>10</sub> , Prophet unggul untuk peramalan jangka panjang.
4.	(Maltare & Vahora, 2023)	Peramalan AQI menggunakan <i>machine learning</i>	<i>Support Vector Regression</i> (SVR) dengan kernel <i>Radial Basis Function</i> (RBF), dan LSTM	Data kualitas udara dari Ahmedabad, Gujarat, India.	SVR (RBF) unggul dibanding SARIMA dan LSTM berdasarkan R <sup>2</sup> & RMSE
5.	(Bahadure <i>et al.</i> , 2025)	Klasifikasi dan peramalan AQI menggunakan SVR dan SARIMA	SVR, SARIMA	Data kualitas udara dari kota-kota di India (2018–2023)	Akurasi 97%, lebih baik dari ARIMA klasik (94,1%)
6.	(Abdelmalik <i>et al.</i> , 2025)	Peramalan AQI menggunakan <i>machine learning</i>	<i>Gaussian Process Regression</i> (GPR), <i>Ensemble Regression</i> (ER), SVM,	Data kualitas udara Gazipur, Bangladesh (2022)	GPR, ER, SVM, dan RT akurat (>96%), GPR terbaik dengan

			<i>Regression Tree (RT)</i> , dan <i>Kernel Approximation Regression (KAR)</i>		RMSE terendah.
7.	(B. Liu <i>et al.</i> , 2021)	Peramalan kualitas udara di Nanjing menggunakan <i>Principal Component Regression (PCR)</i> – <i>SVR</i> – <i>Autoregressive Moving Average (ARMA)</i>	PCR, SVR, ARMA	Data kualitas udara Nanjing (2018–2019)	Model gabungan meningkatkan akurasi hingga 93,2%
8.	(Castelli <i>et al.</i> , 2020)	Peramalan kualitas udara menggunakan SVR	SVR dengan kernel <i>Radial Basis Function (RBF)</i>	Data kualitas udara di California, termasuk polutan seperti CO, SO <sub>2</sub> , NO <sub>2</sub> , O <sub>3</sub> , dan PM2.5.	SVR dengan kernel RBF berhasil meramalkan konsentrasi polutan dan AQI dengan akurasi 94,1% pada data validasi yang tidak terlihat
9.	(Trong <i>et al.</i> , 2026)	Sistem pemantauan dan peramalan kualitas udara <i>real-time</i> berbasis	IoT + sensor canggih + <i>fog computing</i> + model ML (Transformer, CNN, RF, SVM, LSTM)	Sistem kualitas udara portabel untuk <i>monitoring &amp; forecasting</i> di lingkungan urban	Sistem meningkatkan akurasi, fleksibilitas, dan efisiensi dibanding

		IoT dengan ML & <i>fog computing</i> .			metode tradisional
10.	(Md. M. Rahman <i>et al.</i> , 2024)	Peramalan kualitas udara global memakai model <i>machine learning</i> termasuk <i>Random Forest</i> , <i>Decision Tree</i> , SVM, dan lainnya dengan antarmuka web <i>real-time</i>	Beragam algoritma ML ( <i>Random Forest</i> , <i>Decision Tree</i> , SVM, Linear SVC, KNN, <i>Logistic Regression</i> , <i>Naïve Bayes</i> ) dilatih pada data polutan CO, O <sub>3</sub> , NO <sub>2</sub> , PM2.5 dari 23.463 kota; sistem web Django dibuat untuk <i>monitoring</i>	Dataset kualitas udara global dari ribuan kota di seluruh dunia, fokus pada parameter utama polusi udara	<i>Random Forest</i> dan <i>Decision Tree</i> mencapai akurasi sangat tinggi pada dataset yang digunakan untuk peramalan kategori kualitas udara <i>real-time</i>
11.	(Z. Liu <i>et al.</i> , 2024)	Peramalan konsentrasi PM2.5 dengan model <i>hybrid</i> optimisasi	mWOA-SVR ( <i>Modified Whale Optimization</i> + SVR)	Data PM2.5, SO <sub>2</sub> , CO + meteorologi dari Nanchang, China	mWOA-SVR meningkatkan akurasi peramalan PM2.5 dibanding SVR standar
12.	(Kalantari <i>et al.</i> , 2024)	Perbandingan model <i>machine learning</i> & <i>deep learning</i> untuk peramalan AQI	SVM, RF, KNN, ANN, LSTM, GRU, RNN, CNN	Data harian PM2.5 & meteorologi di Zabol, Iran	<i>Deep learning</i> unggul dalam beberapa metrik, semua model mampu meramalkan AQI dengan tingkat

					AUC tinggi
13.	(Alfasanah <i>et al.</i> , 2025)	Pemantauan polusi udara PM2.5 menggunakan XGBoost dan SVR	XGBoost, SVR, dan <i>Exponentially Weighted Moving Average</i> (EWMA) <i>Charts</i>	Data kualitas udara PM2.5 dari Jakarta	SVR efektif untuk autokorelasi, dan kombinasi XGBoost-EWMA mendeteksi sedikit OOC, meningkatkan akurasi dan mengurangi alarm palsu
14.	(Batool Awan & Ali, 2025)	Monitoring dan peramalan karakteristik PM2.5	Standardized PM2.5 Index (SPM2.5I), <i>Probabilistic Forecasting</i>	Data PM2.5 dari berbagai kota Pakistan	Pendekatan indeks standar efektif untuk memodelkan dan meramalkan PM2.5
15.	(Fatima <i>et al.</i> , 2025)	AQI <i>time-series forecasting</i>	Kalman Attention Bi-GRU + Chi-Square Divergence, ARIMA imputasi	Hourly pollutant data (EPA, US: CO, NO <sub>2</sub> , SO <sub>2</sub> , O <sub>3</sub> , PM <sub>10</sub> , PM2.5)	Model LSTM/CNN-LSTM (R <sup>2</sup> ~0.9679; MSE & MAE rendah)

Berdasarkan Tabel 2.1, mayoritas penelitian menunjukkan bahwa SVR, khususnya dengan kernel RBF, efektif untuk peramalan kualitas udara jangka pendek. Beberapa studi juga memperlihatkan bahwa optimasi parameter menggunakan algoritma metaheuristik dapat meningkatkan performa model.

Meskipun demikian, sebagian besar penelitian masih menggunakan dataset statis, berfokus pada satu wilayah, dan belum banyak mengintegrasikan proses pengambilan data, praproses, pelatihan model, serta penyajian hasil dalam satu sistem operasional. Oleh karena itu, penelitian ini mengembangkan sistem peramalan PM2.5 berbasis SVR yang memanfaatkan WAQI API dan diimplementasikan dalam sistem informasi berbasis web.

## 2.2 Pengantar Kualitas Udara

Kualitas udara merupakan kondisi udara ambien yang memengaruhi kesehatan, lingkungan, dan aktivitas masyarakat (Haya et al., 2025). Penurunan kualitas udara umumnya disebabkan oleh meningkatnya konsentrasi polutan di atmosfer, terutama Particulate Matter berdiameter  $\leq 2,5$  mikrometer (PM2.5), yang berisiko tinggi terhadap kesehatan manusia. Paparan PM2.5 dalam jangka panjang dapat meningkatkan risiko gangguan pernapasan, penyakit kardiovaskular, dan kematian prematur (Haya et al., 2025). Oleh karena itu, kualitas udara menjadi indikator penting dalam menilai tingkat keamanan lingkungan. Dalam perspektif Islam, kualitas lingkungan juga berkaitan dengan prinsip keseimbangan ciptaan Allah SWT. Hal ini sebagaimana dijelaskan dalam Al-Qur'an Surah Al-Hijr ayat 19:

وَالْأَرْضَ مَدَدْنَاهَا وَأَلْقَيْنَا فِيهَا رُوسِيَ وَأَنْبَتْنَا فِيهَا مِنْ كُلِّ شَيْءٍ مَّوْزُونٍ

“Dan Kami telah menghamparkan bumi dan menjadikan padanya gunung-gunung dan Kami tumbuhkan padanya segala sesuatu menurut ukuran” (QS. Al-Hijr:19).

Ayat tersebut menegaskan bahwa bumi diciptakan dalam keadaan teratur dan seimbang. Ibnu Katsir menjelaskan bahwa frasa *menurut ukuran* menunjukkan

bahwa setiap ciptaan memiliki kadar dan fungsi yang sesuai dalam tatanan kehidupan (Shaykh & Ghoffar, 2006). Dengan demikian, pencemaran udara dapat dipandang sebagai bentuk gangguan terhadap keseimbangan lingkungan, sehingga pemantauan dan peramalan kualitas udara menjadi penting untuk mendukung keberlanjutan kehidupan.

### 2.3 Indeks Kualitas Udara

Indeks Kualitas Udara (*Air Quality Index/AQI*) merupakan indikator kuantitatif yang digunakan untuk menggambarkan tingkat pencemaran udara di suatu wilayah (Umah & Gusmira, 2024). Nilai AQI dihitung berdasarkan konsentrasi beberapa polutan utama di atmosfer, seperti PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, dan O<sub>3</sub>. Indeks ini digunakan untuk menyederhanakan informasi kualitas udara sehingga lebih mudah dipahami oleh masyarakat dan pembuat kebijakan.

Dalam penelitian ini, variabel yang diramalkan adalah konsentrasi PM<sub>2.5</sub>. Nilai PM<sub>2.5</sub> hasil peramalan kemudian digunakan untuk menghitung *Individual Air Quality Index* (IAQI) dan menentukan kategori kualitas udara. Pendekatan ini memungkinkan hasil peramalan tetap dinyatakan dalam satuan konsentrasi polutan, namun dapat diinterpretasikan menggunakan skala indeks kualitas udara. Parameter PM<sub>2.5</sub> dipilih karena memiliki dampak kesehatan yang paling signifikan dibandingkan polutan lainnya (Haya et al., 2025; Syuhada et al., 2023). Skala IAQI untuk PM<sub>2.5</sub> berkisar antara 0 hingga 500, di mana nilai yang lebih tinggi menunjukkan kualitas udara yang semakin buruk serta risiko kesehatan yang semakin meningkat (Bahadure et al., 2025). Secara matematis, perhitungan IAQI



untuk suatu polutan  $p$  dapat dinyatakan dalam bentuk persamaan linier sebagai berikut:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} x (C_p - BP_{Lo}) + I_{Lo} \quad (2.1)$$

Keterangan:

- $I_p$  : Indeks untuk polutan  $p$   
 $C_p$  : Konsentrasi terpotong dari polutan  $p$   
 $BP_{Hi}$  : Titik batas konsentrasi yang lebih besar atau sama dengan  $C_p$   
 $BP_{Lo}$  : Titik batas konsentrasi yang lebih kecil atau sama dengan  $C_p$   
 $I_{Hi}$  : Nilai AQI yang sesuai  $BP_{Hi}$   
 $I_{Lo}$  : Nilai AQI yang sesuai  $BP_{Lo}$

Ringkasan breakpoint US-EPA (U.S. Environmental Protection Agency) untuk PM2.5 yang umum digunakan dalam pemantauan kualitas udara ditunjukkan pada Tabel 2.2 (U.S. Environmental Protection Agency, n.d.).

Tabel 2.2 *Breakpoints* US EPA PM2.5

Konsentrasi PM2.5 24 jam ( $\mu\text{g}/\text{m}^3$ )	AQI	Kategori
0.0–9.0	0–50	Baik
9.1–35.4	51–100	Sedang
35.5–55.4	101–150	Tidak Sehat untuk Kelompok Sensitif
55.5–125.4	151–200	Tidak Sehat
125.5–225.4	201–300	Sangat Tidak Sehat
225.5–500.4	301–500	Berbahaya

Di Indonesia, kualitas udara juga dirangkum melalui Indeks Standar Pencemar Udara (ISPU) yang ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK). Untuk PM2.5, ISPU dihitung dari rata-rata konsentrasi 24 jam menggunakan pemetaan linier terhadap breakpoint tertentu (Kementerian Lingkungan Hidup dan Kehutanan, n.d.).

Tabel 2.3 *Breakpoints* ISPU KLHK PM2.5

Konsentrasi PM2.5 24 jam ( $\mu\text{g}/\text{m}^3$ )	ISPU	Kategori
0.0–15.5	0–50	Baik
15.6–55.4	51–100	Sedang

55.5–150.4	101–200	Tidak Sehat
150.5–250.4	201–300	Sangat Tidak Sehat
250.5–500.0	301–500	Berbahaya

Dalam penelitian ini, nilai PM<sub>2.5</sub> hasil peramalan dikonversi menjadi IAQI dan kemudian dikategorikan berdasarkan standar US-EPA dan ISPU. Pendekatan ini memungkinkan hasil peramalan tidak hanya menunjukkan konsentrasi polutan, tetapi juga memberikan interpretasi kategori kualitas udara yang lebih mudah dipahami. Selain variabel target PM<sub>2.5</sub>, penelitian ini juga menggunakan parameter meteorologi yang dilaporkan oleh WAQI, yaitu suhu udara ( $t$ ), kelembapan relatif ( $h$ ), tekanan udara ( $p$ ), kecepatan angin ( $w$ ), dan titik embun (*dew point*). Parameter meteorologi tersebut berperan sebagai variabel independen yang dapat memengaruhi fluktuasi konsentrasi PM<sub>2.5</sub> di atmosfer (Rajesh *et al.*, 2025). Misalnya, suhu dan kelembapan memengaruhi pembentukan partikel sekunder, tekanan udara berkaitan dengan dinamika massa udara, kecepatan angin memengaruhi dispersi polutan, sedangkan titik embun berhubungan dengan proses kondensasi yang dapat meningkatkan konsentrasi partikulat (R. A. Rahman *et al.*, 2024).

Dengan memanfaatkan konsentrasi PM<sub>2.5</sub> sebagai target peramalan serta dukungan parameter meteorologi, model yang dikembangkan dapat merepresentasikan dinamika kualitas udara secara lebih komprehensif. Hasil peramalan tersebut kemudian dapat digunakan untuk menghitung IAQI dan menentukan kategori kualitas udara sehingga dapat dimanfaatkan sebagai informasi peringatan dini dan dasar pengambilan keputusan dalam pengelolaan kualitas udara (Rajesh *et al.*, 2025).

## 2.4 WAQI API sebagai Sumber Data Kualitas Udara

*World Air Quality Index* (WAQI) merupakan platform global yang menyediakan informasi kualitas udara secara berkala dengan cakupan lebih dari 130 negara, termasuk Indonesia (The World Air Quality Index, n.d.). Data yang disediakan oleh WAQI merupakan hasil agregasi dari berbagai sumber pemantauan kualitas udara, termasuk stasiun pemerintah, lembaga penelitian, serta mitra lokal. Pada beberapa lokasi di Indonesia, metadata atribusi menunjukkan keterkaitan dengan lembaga resmi seperti Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) (The World Air Quality Index, n.d.). Salah satu keunggulan utama WAQI adalah ketersediaannya dalam bentuk *Application Programming Interface* (API). Melalui API tersebut, pengguna dapat mengakses data kualitas udara secara berkala tanpa perlu membangun infrastruktur sensor pemantauan sendiri (Sarmadi *et al.*, 2021). Data yang tersedia mencakup berbagai parameter kualitas udara dan meteorologi yang dapat dimanfaatkan dalam analisis lingkungan maupun pengembangan sistem peramalan.

Sejumlah penelitian menunjukkan bahwa WAQI API dapat dimanfaatkan secara efektif untuk membangun sistem pemantauan dan peramalan kualitas udara berbasis data historis. Penelitian Joshi *et al.* (2025) menunjukkan bahwa pemanfaatan WAQI API dapat mendukung pengembangan sistem peringatan dini terhadap polusi udara. Selain itu, penggunaan data terbuka dari WAQI juga memungkinkan integrasi dengan *pipeline machine learning* untuk pemodelan lingkungan berbasis data deret waktu (Rajesh *et al.*, 2025). Dengan cakupan wilayah yang luas, sifat data yang terbuka, dan kemudahan integrasi dengan

*pipeline machine learning*, WAQI API menjadi sumber data yang relevan untuk penelitian ini.

## 2.5 Support Vector Regression (SVR) dalam Peramalan Kualitas Udara

*Support Vector Regression* (SVR) merupakan salah satu varian dari *Support Vector Machine* (SVM) yang digunakan untuk menyelesaikan masalah regresi (Shafii *et al.*, 2020). Prinsip dasar SVR adalah membangun fungsi regresi yang meminimalkan kesalahan peramalan dengan tetap mempertahankan margin maksimum antara data dan fungsi peramalan tersebut (Castelli *et al.*, 2020). Ini menjadikan SVR tahan terhadap fluktuasi data yang tinggi, terutama jika dibandingkan dengan metode regresi linier konvensional (Shafii *et al.*, 2020). Secara matematis, SVR berusaha menemukan fungsi regresi dalam bentuk

$$f(x) = \langle w, x \rangle + b \quad (2.2)$$

Keterangan:

$f(x)$  : Fungsi yang bergantung pada  $x$   
 $w$  : Bobot (*weight*)  
 $x$  : Input atau fitur  
 $\langle w, x \rangle$  : hasil perkalian titik (dot product) antara  $w$  dan  $x$   
 $b$  : bias

Dalam praktiknya, SVR sering menggunakan fungsi kernel untuk menangani hubungan *non-linear* antara variabel input dan target. Salah satu kernel yang paling umum digunakan adalah *Radial Basis Function* (RBF). Kernel ini memungkinkan data dipetakan ke ruang berdimensi lebih tinggi sehingga pola kompleks dalam data dapat dipelajari dengan lebih baik (Maltare & Vahora, 2023). Fungsi kernel  $K(x, x')$  digunakan untuk mengukur kemiripan antara dua titik data. Salah satu parameter penting dalam kernel RBF adalah parameter  $\gamma$  (gamma), yang

mengontrol jangkauan pengaruh suatu titik data terhadap titik lainnya. Nilai  $\gamma$  yang kecil menghasilkan pengaruh yang lebih luas, sedangkan nilai  $\gamma$  yang besar membuat model lebih sensitif terhadap variasi lokal pada data (Houssein *et al.*, 2025).

Sebelum diterapkan pada model SVR, data biasanya melalui tahap praproses. Salah satu langkah yang umum dilakukan adalah standardisasi fitur menggunakan *StandardScaler*, yang mentransformasikan setiap variabel sehingga memiliki rata-rata nol dan standar deviasi satu. Proses ini bertujuan untuk menyamakan skala antar fitur agar tidak terjadi dominasi variabel tertentu dalam proses pelatihan model (Muhaimin *et al.*, 2024). Selain itu, seleksi fitur juga dapat dilakukan untuk memilih variabel yang paling relevan terhadap target peramalan. Salah satu metode seleksi fitur yang sering digunakan adalah *SelectKBest*, yaitu metode univariat yang menilai hubungan antara fitur dan target berdasarkan skor statistik tertentu. Metode ini relatif sederhana serta efisien secara komputasi dalam mengidentifikasi fitur yang memiliki kontribusi paling signifikan terhadap model (Muhaimin *et al.*, 2024; Wong *et al.*, 2025).

Berbagai penelitian menunjukkan bahwa SVR mampu memberikan performa yang baik dalam peramalan kualitas udara (Bahadure *et al.*, 2025). Maltare & Vahora (2023) juga menunjukkan bahwa SVR dengan kernel RBF mampu meningkatkan akurasi peramalan indeks kualitas udara dibandingkan model statistik konvensional. Selain itu, kombinasi SVR dengan teknik optimasi seperti *Whale Optimization Algorithm* (WOA) maupun *Harris Hawks Optimization* (HHO) juga terbukti mampu meningkatkan performa model dalam meramalkan

konsentrasi PM2.5 (Bahadure *et al.*, 2025; Houssein *et al.*, 2025; Z. Liu *et al.*, 2024). Oleh karena itu, SVR dengan kernel RBF dipandang sebagai pendekatan yang tepat untuk membangun sistem peramalan kualitas udara berbasis data deret waktu. Penggunaan metode peramalan berbasis data kuantitatif dan proses verifikasi parameter dalam SVR menunjukkan pentingnya landasan ilmiah yang valid dalam setiap pengambilan keputusan. Prinsip kehati-hatian dan keharusan berlandaskan ilmu tersebut sejalan dengan nilai yang terkandung dalam Al-Qur'an Surat Al-Isra ayat 36:

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهِ عِلْمٌ ۚ إِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ أُولَٰئِكَ كَانَ عَنْهُ مَسْئُولًا

*“Dan janganlah kamu mengikuti apa yang kamu tidak mempunyai pengetahuan tentangnya. Sesungguhnya pendengaran, penglihatan dan hati, semuanya itu akan diminta pertanggungjawaban” (QS. Al-Isra:36).*

Ayat ini menegaskan bahwa Allah SWT melarang manusia mengikuti, mengatakan, atau melakukan sesuatu tanpa didasari ilmu dan pengetahuan yang benar, serta menegaskan bahwa pendengaran, penglihatan, dan hati akan dimintai pertanggungjawaban. Ulama tafsir seperti Ibnu Katsir menjelaskan bahwa ayat ini melarang seseorang mengikuti atau mengatakan sesuatu tanpa didasari pengetahuan yang benar, seperti mengaku melihat, mendengar, atau mengetahui sesuatu padahal tidak demikian, serta memperingatkan bahwa pendengaran, penglihatan, dan hati akan dimintai pertanggungjawaban atas penggunaannya (Shaykh & Ghoffar, 2006). Makna ini menunjukkan pentingnya verifikasi, kehati-hatian, dan tanggung jawab moral dalam setiap ucapan dan tindakan dalam konteks modern, prinsip tersebut relevan dengan keharusan menggunakan data yang valid dan metode yang dapat dipertanggungjawabkan, termasuk dalam penelitian dan peramalan kualitas udara,

agar keputusan yang diambil tidak didasarkan pada asumsi semata, melainkan pada pengetahuan yang sah.

## 2.6 Transformasi *Yeo-Johnson* pada Variabel Target

Transformasi *Yeo-Johnson* merupakan metode transformasi daya (*power transformation*) yang digunakan untuk memperbaiki karakteristik distribusi data, terutama pada data yang memiliki distribusi tidak simetris (*skewed*) atau varians yang tidak stabil (Riani *et al.*, 2023). Metode ini merupakan pengembangan dari transformasi *Box-Cox* dan memiliki keunggulan karena dapat diterapkan pada data yang mengandung nilai nol maupun nilai negatif, sehingga lebih fleksibel dibandingkan beberapa metode transformasi lainnya (Riani *et al.*, 2023). Dalam penelitian ini, transformasi *Yeo-Johnson* diterapkan pada variabel target, yaitu konsentrasi PM2.5.

Pada studi peramalan kualitas udara, distribusi konsentrasi PM2.5 sering menunjukkan pola yang tidak simetris serta dipengaruhi oleh nilai ekstrem. Kondisi ini menyebabkan karakteristik data target menjadi kurang ideal untuk proses pembelajaran model. Penelitian sebelumnya menunjukkan bahwa data polutan udara memiliki pola yang kompleks, fluktuatif, dan tidak selalu mengikuti distribusi normal (Castelli *et al.*, 2020). Oleh karena itu, transformasi pada variabel target dapat membantu mengurangi tingkat *skewness*, menstabilkan varians, serta meningkatkan kemampuan model dalam mempelajari hubungan antara variabel input dan target (Riani *et al.*, 2023). Secara matematis, transformasi *Yeo-Johnson* dapat dinyatakan sebagai berikut:

$$T(y; \lambda) = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda}, & \text{Jika } y \geq 0, \lambda \neq 0 \\ \log(y+1), & \text{Jika } y \geq 0, \lambda = 0 \\ -\frac{(-y+1)^{2-\lambda} - 1}{2-\lambda}, & \text{Jika } y < 0, \lambda \neq 2 \\ -\log(1-y), & \text{Jika } y < 0, \lambda = 2 \end{cases} \quad (2.3)$$

Keterangan:

- $T(y; \lambda)$  : nilai hasil transformasi *Yeo-Johnson*  
 $y$  : nilai asli variabel target  
 $\lambda$  : parameter transformasi yang diestimasi secara otomatis  
 $y \geq 0$  : kondisi untuk data bernilai nol atau positif  
 $y < 0$  : kondisi untuk data bernilai negatif

Parameter  $\lambda$  diestimasi secara otomatis untuk memperoleh bentuk transformasi yang paling sesuai dengan karakteristik data (Riani *et al.*, 2023). Setelah transformasi dilakukan, nilai target kemudian dinormalisasi melalui proses standardisasi sehingga memiliki rata-rata nol dan standar deviasi satu. Tahap ini membantu meningkatkan stabilitas numerik serta mempercepat proses konvergensi pada model SVR. Dalam penelitian ini, transformasi *Yeo-Johnson* hanya diterapkan pada variabel target PM2.5. Meskipun nilai PM2.5 umumnya tidak bernilai negatif, metode ini tetap relevan karena mampu menangani nilai nol serta membantu menghasilkan distribusi target yang lebih simetris.

Pada pemodelan *Support Vector Regression*, transformasi target tidak digunakan untuk memenuhi asumsi normalitas secara ketat, melainkan untuk memperbaiki karakteristik distribusi data agar proses pembelajaran model menjadi lebih stabil (Castelli *et al.*, 2020). Dengan berkurangnya keasimetrisan distribusi serta pengaruh nilai ekstrem, model diharapkan mampu mempelajari pola hubungan antara variabel input dan target secara lebih efektif. Setelah proses peramalan dilakukan, hasil peramalan kemudian dikembalikan ke skala asli



konsentrasi PM2.5 melalui proses *inverse transform* sehingga tetap dapat diinterpretasikan dalam satuan yang sebenarnya.

## 2.7 Validasi dan Evaluasi Model Peramalan

Validasi merupakan tahap penting dalam pengembangan model peramalan karena menentukan sejauh mana model mampu merepresentasikan pola data secara akurat serta dapat digeneralisasikan pada data yang belum pernah dilihat sebelumnya (Asrul & Soebroto, 2024). Dalam pemodelan deret waktu (*time series*), proses validasi perlu mempertimbangkan urutan kronologis data agar evaluasi performa tetap sesuai dengan karakteristik temporal yang melekat pada data tersebut (Asrul & Soebroto, 2024). Pendekatan validasi berbasis waktu, seperti *walk-forward validation*, lebih sesuai digunakan pada data deret waktu karena metode tersebut mempertahankan urutan data dan meniru kondisi peramalan pada situasi nyata (Al-Saeedi *et al.*, 2025; Asrul & Soebroto, 2024). Dalam penelitian ini digunakan pendekatan *walk-forward validation*, di mana model dilatih menggunakan data historis dan diuji pada data yang lebih baru secara bertahap. Pendekatan ini memungkinkan evaluasi performa model dilakukan secara realistis tanpa terjadi kebocoran informasi dari masa depan (*data leakage*).

Selain teknik validasi, pemilihan metrik evaluasi juga menjadi aspek krusial dalam menilai kinerja model, beberapa metrik evaluasi yang umum digunakan dalam peramalan kualitas udara adalah *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi ( $R^2$ ) (Maltare & Vahora, 2023). MAE mengukur rata-rata kesalahan absolut sehingga memberikan penilaian yang

lebih *robust* terhadap keberadaan *outlier* (Maltare & Vahora, 2023). MAE dihitung dengan rumus:

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (2.4)$$

Keterangan:

*MAE* : *Mean Absolute Error* (Kesalahan Absolut Rata-rata)  
*n* : Jumlah data atau sampel  
*y<sub>i</sub>* : nilai aktual (nilai yang sesungguhnya) untuk data ke-*i*  
*ŷ<sub>i</sub>* : nilai peramalan (nilai yang diramalkan oleh model) untuk data ke-*i*

Sementara itu, RMSE lebih sensitif terhadap kesalahan besar karena nilai kesalahan dikuadratkan sebelum dihitung rata-ratanya, sehingga cocok digunakan untuk mengidentifikasi seberapa besar deviasi ekstrem dalam peramalan (Maltare & Vahora, 2023). RMSE dihitung dengan rumus:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.5)$$

Keterangan:

*RMSE* : *Root Mean Squared Error* (Akar Kuadrat Kesalahan Rata-rata),  
*n* : Jumlah data atau sampel  
*y<sub>i</sub>* : nilai aktual (nilai yang sesungguhnya) untuk data ke-*i*  
*ŷ<sub>i</sub>* : nilai peramalan (nilai yang diramalkan oleh model) untuk data ke-*i*

Selain MAE dan RMSE, penelitian ini juga menggunakan koefisien determinasi  $R^2$  untuk mengukur kemampuan model dalam menjelaskan variasi data. Nilai  $R^2$  yang mendekati 1 menunjukkan bahwa model mampu merepresentasikan pola data dengan baik, sedangkan nilai negatif mengindikasikan bahwa model memberikan hasil yang lebih buruk dibandingkan pendekatan sederhana menggunakan nilai rata-rata (Chicco *et al.*, 2021). Nilai  $R^2$  dihitung dengan rumus:

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (2.6)$$

Keterangan:

- $R^2$  : Koefisien Determinasi yang mengukur seberapa baik model dalam menjelaskan variasi data  
 $m$  : Jumlah data atau sampel  
 $Y_i$  : nilai aktual (nilai yang sesungguhnya) untuk data ke- $i$   
 $X_i$  : nilai peramalan (nilai yang diramalkan oleh model) untuk data ke- $i$   
 $\bar{Y}$  : nilai rata-rata dari data aktual

Dalam peramalan deret waktu, tidak semua observasi memiliki kontribusi yang sama terhadap proses pembelajaran model. Beberapa penelitian menunjukkan bahwa pemberian bobot yang berbeda pada tiap observasi dapat membantu model menekankan data yang lebih relevan dalam proses peramalan (Xiang *et al.*, 2022). Namun, pendekatan tersebut hanya diuji sebagai eksperimen tambahan dan tidak digunakan pada konfigurasi akhir karena tidak menunjukkan peningkatan performa yang konsisten. Oleh karena itu, evaluasi model dilakukan menggunakan *walk-forward validation* dengan metrik MAE, RMSE, dan  $R^2$  untuk menilai performa model pada data deret waktu kualitas udara.

## **BAB III**

### **DESAIN DAN IMPLEMENTASI**

#### **3.1 Data Preparation**

Penelitian ini menggunakan data *World Air Quality Index* (WAQI) dari 15 stasiun di Indonesia dengan variabel target berupa nilai PM2.5 (pm25) dan variabel independen berupa parameter meteorologi, yaitu suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun. Data dikumpulkan secara berkala dan disimpan dalam basis data MySQL dengan *observed\_at* yang berasal dari atribut *time.s* sebagai waktu pengamatan lokal. Tahap praproses meliputi validasi nilai serta rekayasa fitur *lag*, *rolling window*, *first difference*, *Exponentially Weighted Moving Average* (EWMA), dan fitur musiman, sehingga data siap digunakan untuk pelatihan model SVR dalam meramalkan PM2.5 pada horizon H+1, yaitu satu hari kalender berikutnya. Estimasi harian tersebut diperoleh menggunakan strategi *recursive multi-step forecasting* sebanyak 24 langkah, kemudian hasil peramalan pada 24 langkah tersebut dirata-ratakan untuk merepresentasikan nilai PM2.5 harian. Hasil peramalan PM2.5 kemudian dihitung menjadi IAQI dan dikategorikan menggunakan standar US-EPA dan ISPU.

##### **3.1.1 Sumber dan Variabel**

Penelitian ini memanfaatkan data *World Air Quality Index* (WAQI) sebagai sumber utama. Data dikumpulkan dari 15 wilayah atau stasiun pemantauan di Indonesia (Tabel 3.1). Untuk setiap wilayah, WAQI menyediakan nilai indeks dan parameter meteorologi beserta waktu pengamatan lokal yang tercatat pada atribut

*time.s* (format YYYY-MM-DD HH:MM:SS). Penelitian ini menggunakan waktu tersebut apa adanya tanpa konversi UTC agar interpretasi horizon H+1 tetap sesuai kalender lokal pada setiap wilayah.

Tabel 3.1 Lokasi Stasiun WAQI yang Digunakan dalam Penelitian

No	Wilayah	Kota/Kabupaten	Provinsi
1	Maros	Kab. Maros	Sulawesi Selatan
2	Pangkalanbun	Kab. Kotawaringin Barat	Kalimantan Tengah
3	Sleman	Kab. Sleman	DI Yogyakarta
4	Malang	Kota Malang	Jawa Timur
5	Banjarbaru	Kota Banjar Baru	Kalimantan Selatan
6	Medan	Kota Medan	Sumatera Utara
7	Jambi	Kota Jambi	Jambi
8	Talang Betutu	Kota Palembang	Sumatera Selatan
9	Semarang	Kota Semarang	Jawa Tengah
10	Pekanbaru	Kota Pekanbaru	Riau
11	Kototabang	Kabupaten Agam	Sumatera Barat
12	Bengkulu	Kota Bengkulu	Bengkulu
13	Samarinda	Kota Samarinda	Kalimantan Timur
14	Kemayoran	Jakarta Pusat	DKI Jakarta
15	Sorong	Sorong	Papua Barat

Variabel target dalam penelitian ini adalah PM2.5 (pm25) yang diperoleh dari data WAQI. Variabel independen diambil dari objek *data.iaqi*, yang meliputi suhu (t), kelembapan relatif (h), tekanan udara (p), kecepatan angin (w), dan titik embun (dew). Nilai PM2.5 hasil peramalan kemudian digunakan untuk menghitung IAQI dan menentukan kategori kualitas udara berdasarkan standar US-EPA dan ISPU.

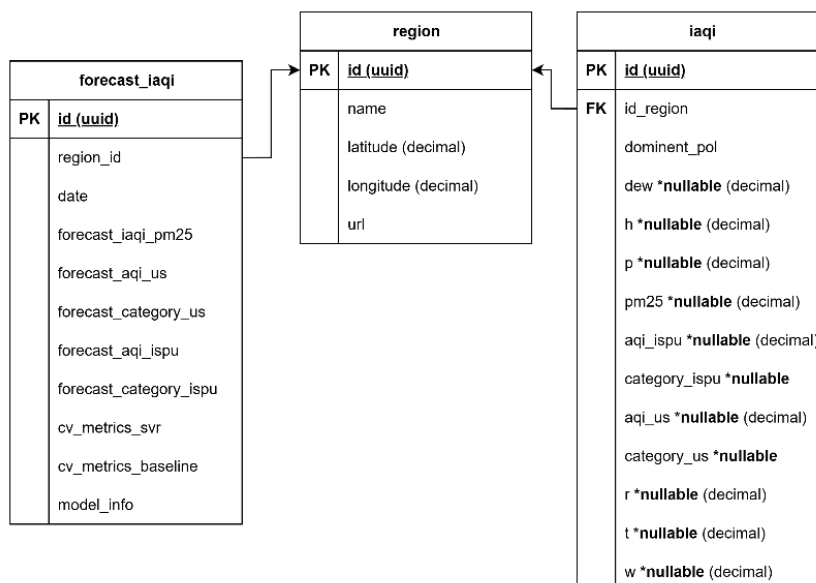
### 3.1.2 Akuisisi Data

Proses pengumpulan data dilakukan secara otomatis menggunakan Laravel melalui perintah `app:sync-air-quality` yang dijalankan setiap hari. Sistem mengirimkan permintaan *Hypertext Transfer Protocol* (HTTP) GET ke *endpoint*

*feed* WAQI untuk masing-masing wilayah. Untuk menjaga keberlanjutan akses dan menghindari batasan permintaan, sistem memanfaatkan beberapa token API. Proses akuisisi juga menerapkan mekanisme *timeout* dan *retry*, serta mencatat *log* keberhasilan maupun kegagalan pengambilan data. Selain itu, jeda singkat antarwilayah diterapkan sebagai bentuk etika pemanggilan layanan API. Variabel yang diambil dari WAQI meliputi *pm25*, *t*, *h*, *p*, *w*, *dew*, serta waktu pengamatan lokal dari *time.s*. Nilai *time.s* langsung digunakan sebagai *observed\_at* tanpa konversi zona waktu, sehingga konsistensi temporal antarwilayah tetap terjaga. Sebelum data disimpan ke MySQL, sistem melakukan deduplikasi berdasarkan pasangan *region\_id* dan *observed\_at* untuk mencegah pencatatan ganda. Pendekatan ini memastikan bahwa deret waktu PM2.5 dan parameter meteorologi bersifat konsisten dan mutakhir untuk setiap wilayah kajian.

### 3.1.3 Penyimpanan Data

Data hasil akuisisi disimpan dalam basis data MySQL dengan rancangan skema yang memisahkan metadata wilayah, deret waktu pengamatan, dan keluaran peramalan. Rancangan *Entity Relationship Diagram* (ERD) pada Gambar 3.1.



Gambar 3.1 Diagram Hubungan Entitas Basis Data

Sistem menyimpan informasi wilayah pemantauan pada tabel *regions* yang memuat identitas stasiun, nama kota, provinsi, koordinat geografis, serta URL *feed* WAQI. Sementara itu, tabel *iaqi* digunakan untuk menyimpan data pengamatan kualitas udara pada setiap wilayah dan waktu tertentu, yang mencakup *region\_id*, *observed\_at*, nilai PM2.5 (pm25), serta parameter meteorologi pendukung seperti suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun. Selain itu, sistem juga menyimpan kategori indeks kualitas udara berdasarkan standar ISPU dan US-EPA.

Untuk menjaga konsistensi deret waktu, sistem menetapkan indeks unik gabungan pada pasangan (*region\_id*, *observed\_at*), sehingga proses penyimpanan bersifat idempoten dan mencegah duplikasi data. Keluaran peramalan satu hari ke depan (H+1) disimpan pada tabel *forecast\_iaqi* yang memuat tanggal peramalan lokal, nilai PM2.5 hasil peramalan, nilai IAQI hasil perhitungan, kategori kualitas udara, serta ringkasan metrik evaluasi dan informasi model dalam format JSON.

Seluruh proses penyimpanan mengikuti waktu lokal WAQI sebagaimana tercatat pada *time.s*, sehingga alur data tetap konsisten dari tahap akuisisi hingga tahap peramalan.

### 3.1.4 Praproses

Tahap praproses dilakukan untuk memastikan mutu deret waktu sebelum memasuki proses pemodelan. Sistem memproses data pada setiap wilayah secara terpisah dengan tetap mempertahankan urutan kronologis berdasarkan waktu pengamatan lokal *observed\_at* yang berasal dari WAQI. Seluruh rekaman diurutkan sesuai cap waktu tersebut, kemudian dilakukan deduplikasi berdasarkan pasangan *region\_id* dan *observed\_at* agar hanya terdapat satu observasi untuk setiap waktu pengamatan.

Selanjutnya, sistem memastikan seluruh variabel utama berada dalam format numerik dan menghapus rekaman yang tidak memiliki nilai target pm25 karena tidak dapat digunakan dalam pelatihan maupun peramalan. Selain itu, sistem memvalidasi kontinuitas deret waktu meskipun interval antar pengamatan tidak selalu seragam karena mengikuti ketersediaan data dari sumber WAQI. Untuk menjaga keterkinian data, periode pengamatan juga dibatasi sesuai rentang data yang tersedia pada masing-masing wilayah. Penelitian ini menetapkan bahwa setiap wilayah harus memiliki setidaknya seribu sampel setelah rekayasa fitur agar proses pelatihan model tetap layak dilakukan. Dengan demikian, tahap praproses menghasilkan deret waktu lokal yang bersih, konsisten, dan siap untuk memasuki tahap pemodelan.



### 3.1.5 Rekayasa Fitur

Setelah deret waktu bersih dan lengkap, tahap berikutnya adalah rekayasa fitur untuk mengubah data deret waktu menjadi format *supervised learning* dengan horizon peramalan satu hari ke depan (H+1). Fitur yang dibangun bertujuan untuk menangkap pola historis, tren lokal, serta karakteristik musiman yang memengaruhi fluktuasi konsentrasi PM2.5.

Sistem membentuk fitur *lag* untuk variabel target maupun variabel independen meteorologi, yaitu  $pm25_{t-L}$  serta  $x_{t-L}$  dengan  $x$  mencakup  $t$ ,  $h$ ,  $p$ ,  $w$ , dan  $dew$ . Nilai *lag* merepresentasikan kondisi pada waktu sebelumnya, misalnya,  $pm25\_lag1$  menunjukkan nilai PM2.5 pada data sebelumnya. Selain *lag*, sistem menghitung fitur *rolling window* untuk menangkap tren lokal melalui statistik bergerak dalam jendela waktu tertentu. Fitur *rolling mean* pada variabel target dan meteorologi digunakan untuk menangkap tren lokal dan menstabilkan fluktuasi jangka pendek.

Untuk menangkap pola musiman, sistem menambahkan fitur hari-dalam-pekan yang direpresentasikan dalam bentuk sinus dan kosinus agar pola periodik dapat dipelajari model secara lebih efektif. Sistem juga menambahkan fitur *first difference* pada PM2.5 untuk merepresentasikan perubahan laju konsentrasi antar waktu, sehingga model dapat menangkap dinamika kenaikan atau penurunan polusi. Selain itu, *Exponentially Weighted Moving Average* (EWMA) diterapkan pada variabel target dan variabel independen meteorologi untuk memberikan bobot lebih besar pada observasi terbaru, sehingga model lebih adaptif terhadap perubahan jangka pendek.

## 3.2 Alur dan Desain Sistem

Sistem yang dibangun dalam penelitian ini dirancang untuk mengintegrasikan proses akuisisi data, pemodelan peramalan, dan penyimpanan hasil secara terpusat. Laravel digunakan sebagai komponen utama untuk sinkronisasi data dan orkestrasi proses, Flask berperan sebagai layanan pemodelan berbasis *Support Vector Regression* (SVR), sedangkan MySQL digunakan sebagai basis data untuk menyimpan deret waktu kualitas udara dan hasil peramalan. Seluruh proses dirancang agar aliran data berjalan konsisten dari tahap pengambilan data hingga penyajian hasil pada sistem informasi berbasis web.

Secara umum, alur sistem dimulai dari pengambilan data kualitas udara dan parameter meteorologi dari WAQI oleh Laravel, kemudian data disimpan ke dalam MySQL dan diteruskan ke Flask untuk proses pra-proses, rekayasa fitur, pelatihan model, serta peramalan PM<sub>2.5</sub> satu hari ke depan (H+1). Hasil peramalan yang diperoleh selanjutnya dihitung menjadi IAQI, dikategorikan berdasarkan standar US-EPA dan ISPU, lalu disimpan kembali ke basis data untuk ditampilkan pada aplikasi web. Arsitektur dan alur kerja sistem secara lebih rinci dijelaskan pada subbab berikutnya.

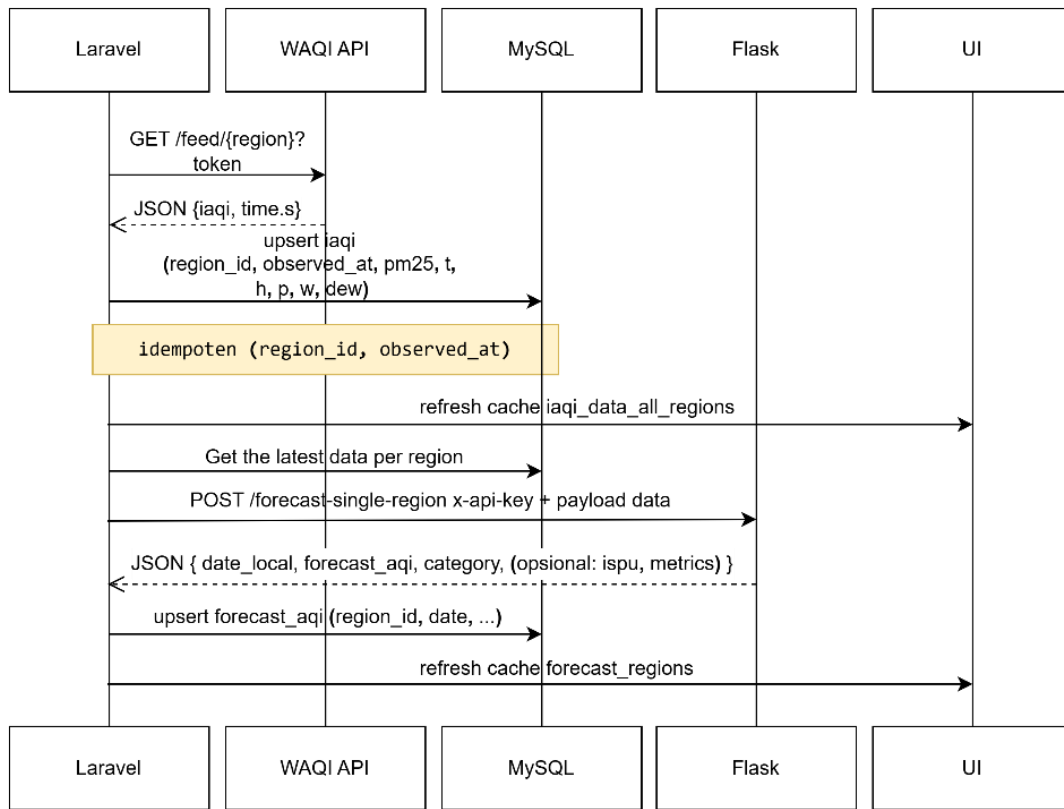
### 3.2.1 Arsitektur *End-to-End*

Arsitektur sistem peramalan kualitas udara yang dibangun menggunakan tiga komponen utama, yaitu Laravel sebagai orkestrator dan lapisan aplikasi, Flask sebagai layanan peramalan, serta MySQL sebagai basis data operasional. Sistem dirancang untuk menjalankan *pipeline end-to-end* mulai dari sinkronisasi data

mentah, pembentukan deret waktu historis, proses peramalan satu hari ke depan (H+1), hingga penyajian hasil pada antarmuka web secara konsisten.

Laravel berperan sebagai pengelola proses sinkronisasi data PM2.5 dan variabel meteorologi dari API *World Air Quality Index* (WAQI). Proses pengambilan dilakukan secara berkala melalui *command scheduler*. Untuk menjaga reliabilitas, pengambilan data dilakukan per wilayah secara bertahap, sehingga beban API lebih terkendali dan risiko kegagalan tidak berdampak pada keseluruhan wilayah sekaligus. Data yang diterima dari WAQI kemudian dipetakan ke skema internal, dinormalisasi sesuai atribut yang dibutuhkan, dan disimpan ke MySQL sebagai data pengamatan. Penyimpanan memanfaatkan kombinasi *region\_id* dan *observed\_at* sebagai kunci unik sehingga setiap pengamatan hanya tercatat satu kali. Mekanisme ini memastikan deret waktu tetap konsisten, mencegah duplikasi, dan memudahkan penarikan data historis berdasarkan urutan waktu pengamatan lokal.

Sistem memastikan bahwa data yang digunakan untuk peramalan jangka pendek memiliki horizon H+1, yaitu nilai yang merepresentasikan kondisi pada hari berikutnya. Dengan demikian, dataset historis yang dibentuk selalu mengikuti struktur sebab akibat yang benar, yaitu variabel penjelas dan target hanya berasal dari waktu  $t$  dan waktu sebelumnya, sedangkan label peramalan mengacu pada waktu berikutnya.



Gambar 3.2 Arsitektur *end-to-end* sistem peramalan berbasis SVR

Gambar 3.2 menggambarkan arsitektur *end-to-end* sistem peramalan PM2.5 berbasis *Support Vector Regression* (SVR). Diagram tersebut memperlihatkan aliran data dari WAQI menuju MySQL melalui Laravel, kemudian diteruskan ke Flask untuk proses peramalan, dan akhirnya dikembalikan lagi ke MySQL untuk ditampilkan pada aplikasi web.

Pada tahap peramalan, Laravel menyiapkan *payload* berupa data historis terbaru untuk setiap wilayah dalam bentuk deret waktu terurut secara kronologis. *Payload* ini mencakup variabel target PM2.5 serta variabel meteorologi sebagai variabel independen, seperti suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun. Data historis dikirimkan ke Flask melalui *endpoint* layanan peramalan secara periodik setelah sinkronisasi data mentah selesai dijalankan.

Urutan eksekusi ini penting agar proses peramalan selalu menggunakan histori yang paling mutakhir dan tidak memproses data yang belum lengkap.

Di sisi Flask, data diterima lalu diproses sebagai deret waktu berdasarkan *observed\_at*. Tahap praproses mencakup pemeriksaan kualitas data, serta penyeragaman format dan urutan waktu. Setelah itu, Flask membangun representasi fitur yang relevan untuk SVR. Selanjutnya dilakukan penskalaan fitur agar SVR dengan kernel RBF dapat bekerja stabil, disertai seleksi fitur untuk mempertahankan variabel yang paling informatif dan menekan *noise*.

Seluruh transformasi fitur dan evaluasi dilakukan dengan *walk-forward validation*. Pemisahan data pelatihan dan pengujian tidak dilakukan secara acak, tetapi mengikuti urutan kronologis agar tidak terjadi kebocoran informasi dari masa depan ke masa lalu. Dengan pengaturan ini, model dilatih hanya dari histori yang tersedia pada saat peramalan dilakukan, sehingga simulasi performa model lebih mendekati kondisi operasional nyata.

Model SVR dengan kernel *Radial Basis Function* (RBF) kemudian dilatih untuk menghasilkan peramalan PM2.5 satu hari ke depan (H+1). Proses pelatihan melibatkan pencarian parameter yang terstruktur, misalnya pada parameter C, epsilon, dan gamma, dengan tetap menjaga urutan waktu saat melakukan evaluasi. *Output* layanan Flask mencakup nilai PM2.5 hasil peramalan, yang selanjutnya dihitung menjadi IAQI dan dikategorikan berdasarkan standar AQI US dan ISPU Indonesia, serta ringkasan metrik evaluasi model untuk memantau kualitas peramalan secara periodik.

Setelah menerima hasil dari Flask, Laravel menyimpan *output* peramalan ke MySQL melalui tabel peramalan menggunakan mekanisme *update-or-create* agar setiap wilayah dan horizon peramalan tersimpan konsisten tanpa duplikasi. Untuk meningkatkan performa akses pada aplikasi web, Laravel juga memperbarui *cache* per wilayah dan *cache* agregat harian. Strategi ini membuat halaman daftar dan halaman detail dapat menampilkan data terbaru dengan latensi rendah, namun tetap memiliki fallback ke basis data apabila cache belum tersedia.

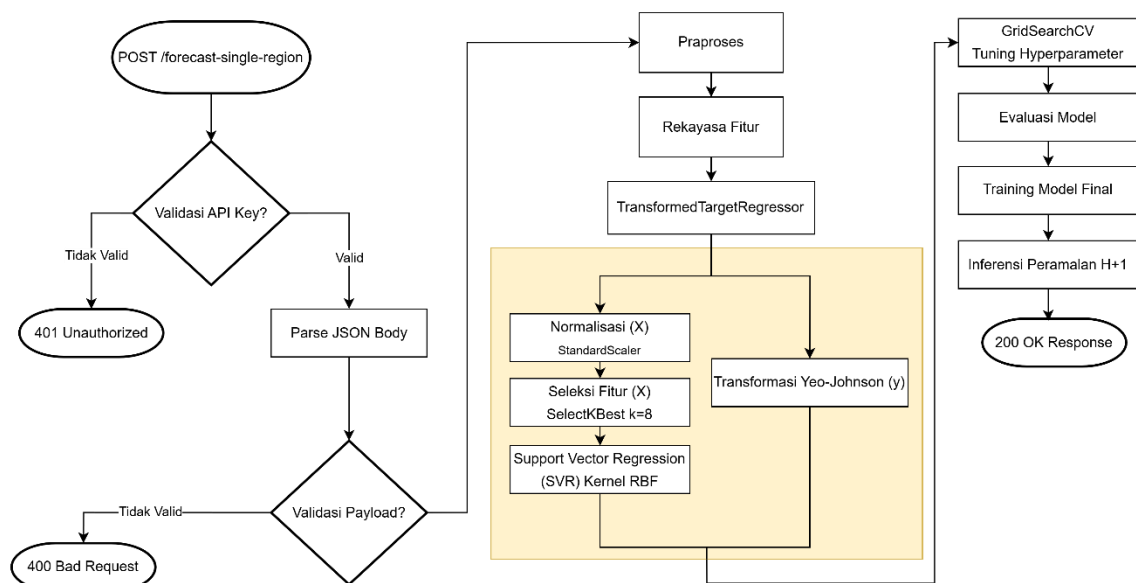
Dengan arsitektur *end-to-end* ini, sistem mampu menjaga kesinambungan aliran data mulai dari akuisisi WAQI, pembentukan deret waktu pengamatan, proses peramalan berbasis SVR, hingga penyajian hasil secara operasional. Desain ini menekankan konsistensi deret waktu, ketertelusuran hasil per wilayah, serta pemisahan yang jelas antara lapisan sinkronisasi data, lapisan peramalan, dan lapisan penyajian informasi.

### 3.2.2 Proses di Flask

Flask berperan sebagai layanan API dan orkestrator proses peramalan yang mengoordinasikan validasi permintaan, pemrosesan data, pelatihan model, serta inferensi pada *endpoint POST /forecast-single-region*. Proses dimulai saat *endpoint* menerima permintaan dari sistem Laravel. Pada tahap awal dilakukan validasi *API key* untuk memastikan bahwa hanya permintaan yang terotorisasi yang dapat diproses. *Endpoint* Flask melakukan validasi autentikasi dan struktur *payload* sebelum data diproses lebih lanjut, sehingga keamanan dan konsistensi input tetap terjaga. Permintaan yang tidak memenuhi kriteria akan dikembalikan sebagai *400*

*Bad Request*. Tahapan ini dirancang untuk menjaga keamanan serta konsistensi data sebelum memasuki proses pemodelan.

Berikutnya, data deret waktu diproses dengan deteksi kolom waktu, pengurutan kronologis, dan pembersihan nilai PM2.5 yang tidak valid sebagai bagian dari praproses. Selanjutnya dilakukan rekayasa fitur agar model dapat menangkap pola fluktuasi kualitas udara yang kompleks dan tidak linear. Horizon peramalan H+1 didefinisikan sebagai satu hari kalender berikutnya yang dihitung dari observasi terakhir yang tersedia pada setiap wilayah. Untuk menghasilkan estimasi harian tersebut, model menerapkan strategi *recursive multi-step forecasting* sebanyak 24 langkah ke depan. Pada setiap langkah, hasil peramalan sebelumnya digunakan kembali sebagai bagian dari input untuk menghasilkan peramalan berikutnya. Selanjutnya, 24 nilai peramalan tersebut dirata-ratakan untuk memperoleh estimasi PM2.5 harian H+1. Gambar 3.3 menunjukkan alur lengkap proses peramalan PM2.5 yang diimplementasikan pada modul Flask.



Gambar 3.3 Alur proses peramalan PM2.5 di Flask

Sebagaimana ditunjukkan pada Gambar 3.3, tahap pemodelan menggunakan pendekatan *TransformedTargetRegressor*, yang memisahkan pengolahan fitur (X) dan transformasi variabel target (y) secara independen. Pada sisi fitur input, data diproses melalui *pipeline* yang terstruktur, dimulai dengan normalisasi menggunakan *StandardScaler* untuk menyamakan skala antar fitur. Kemudian dilakukan seleksi fitur otomatis menggunakan *SelectKBest* dengan  $k=8$  untuk memilih fitur paling informatif dari keseluruhan fitur yang ada. Fitur-fitur terpilih ini kemudian dilatih dengan algoritma *Support Vector Regression* (SVR) dengan kernel RBF. Secara paralel pada sisi target, PM2.5 ditransformasi menggunakan *PowerTransformer* dengan metode *Yeo-Johnson* untuk menstabilkan distribusi dan mengurangi *skewness* sehingga proses pembelajaran model menjadi lebih optimal.

Optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* dengan skema *walk-forward validation* yang menekankan performa pada data terbaru. *Grid* pencarian meliputi parameter C, epsilon, dan gamma dari SVR untuk menemukan kombinasi terbaik. Pendekatan ini bertujuan agar model lebih adaptif terhadap perubahan kondisi atmosfer terkini dan meminimalkan risiko *data leakage*. Setelah kombinasi *hyperparameter* terbaik diperoleh melalui evaluasi model, model final dilatih ulang menggunakan keseluruhan dataset dan digunakan untuk melakukan inferensi peramalan PM2.5 satu hari ke depan (H+1). Hasil peramalan tersebut kemudian dikembalikan ke skala asli PM2.5 melalui *inverse transform*, lalu dihitung menjadi IAQI dan dikonversi ke dalam kategori kualitas udara berdasarkan standar AQI internasional dan ISPU nasional.



Pada tahap akhir, Flask mengembalikan nilai peramalan beserta informasi evaluasi model termasuk metrik evaluasi dan konfigurasi model melalui respons *200 OK* ke sistem Laravel. Laravel kemudian menyimpan hasil tersebut ke dalam tabel peramalan serta mengelolanya melalui mekanisme *cache* harian untuk memastikan stabilitas akses dan efisiensi sistem. Dengan arsitektur ini, proses peramalan berjalan secara terintegrasi mulai dari validasi keamanan, pemrosesan data, optimasi model menggunakan *GridSearchCV*, evaluasi komprehensif, hingga penyajian hasil secara andal dalam implementasi *end-to-end*.

### 3.3 Skenario Pengujian

Skenario pengujian dalam penelitian ini dirancang untuk mengevaluasi kinerja model peramalan PM2.5 secara menyeluruh, baik dari sisi akurasi peramalan maupun keandalan sistem dalam kondisi operasional. Pengujian dilakukan menggunakan data historis deret waktu PM2.5 pada setiap wilayah, yang telah melalui tahapan praproses seperti pengurutan data, rekayasa fitur berbasis *lag*, *rolling window*, *first difference*, *Exponential Weighted Moving Average (EWMA)*, dan fitur musiman harian. Evaluasi kinerja model dilakukan dengan metrik regresi yang umum digunakan dalam peramalan, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan koefisien determinasi ( $R^2$ ), untuk mengukur ketepatan peramalan satu hari ke depan ( $H+1$ ).

Selain evaluasi utama, penelitian ini juga melakukan eksperimen terbatas untuk menilai pengaruh beberapa keputusan desain dalam *pipeline* pemodelan, seperti variasi panjang *lag* dan *rolling window*, penerapan pembobotan waktu eksponensial, serta strategi seleksi fitur. Pengujian operasional turut dilakukan

untuk memastikan sistem berjalan sesuai jadwal, mampu memproses data secara konsisten, serta tetap stabil dalam menghadapi gangguan seperti input tidak valid atau kegagalan komunikasi. Dengan skenario pengujian ini, penelitian memastikan bahwa sistem peramalan yang dibangun tidak hanya akurat secara statistik, tetapi juga andal dalam implementasi nyata.

### 3.3.1. Eksperimen Terbatas

Eksperimen terbatas dilakukan untuk menilai pengaruh beberapa keputusan desain pada *pipeline* pemodelan terhadap akurasi peramalan PM2.5 satu hari ke depan, tanpa mengubah ruang lingkup utama penggunaan model SVR. Seluruh eksperimen dijalankan per wilayah dengan prosedur yang sama seperti implementasi sistem, yaitu praproses data, rekayasa fitur, penambahan fitur musiman hari dalam pekan, normalisasi dan seleksi fitur dalam lipatan validasi, transformasi target, serta evaluasi berbasis waktu secara *walk-forward* dengan pembobotan eksponensial. Seluruh transformasi tetap ditempatkan di dalam alur validasi untuk menjaga validitas evaluasi dan menghindari kebocoran data.

Variasi eksperimen difokuskan pada komponen yang paling berpengaruh secara praktis. Pertama, penelitian menguji kombinasi panjang *lag* dan *rolling window* yang berbeda, mulai dari konfigurasi pendek hingga panjang, untuk menemukan keseimbangan antara kompleksitas fitur dan kestabilan model. Kedua, penelitian mengevaluasi penerapan pembobotan waktu, di mana data terbaru diberi pengaruh lebih besar dibandingkan data lama agar model lebih responsif terhadap dinamika kualitas udara. Ketiga, penelitian membandingkan kondisi tanpa seleksi fitur dengan seleksi fitur aktif menggunakan jumlah fitur terpilih yang berbeda.

Keempat, penelitian mengevaluasi dampak transformasi target monotonik yang bertujuan menstabilkan varians dibandingkan kondisi tanpa transformasi.

Pada setiap skenario, ruang pencarian hiperparameter SVR tetap dipertahankan konsisten agar perbedaan performa dapat diatribusikan pada faktor eksperimen yang diuji. Hasil eksperimen dilaporkan dalam bentuk rata-rata MAE, RMSE, dan  $R^2$ , baik secara per wilayah maupun ringkasan lintas wilayah. Wilayah dengan jumlah sampel yang tidak memenuhi ambang minimal setelah rekayasa fitur dikeluarkan dari eksperimen agar interpretasi hasil tetap valid, sementara *baseline persistence* tetap digunakan sebagai konteks pembandingan.

### 3.3.2. Evaluasi Model

Evaluasi kinerja model dalam penelitian ini dilakukan terhadap hasil peramalan PM2.5 menggunakan metrik regresi utama, yaitu  $R^2$ , MAE, dan RMSE. Nilai PM2.5 hasil peramalan kemudian dihitung menjadi IAQI dan dikategorikan untuk keperluan interpretasi kualitas udara. Koefisien determinasi  $R^2$  digunakan untuk mengukur proporsi variasi data target yang dapat dijelaskan oleh model. Nilai  $R^2$  yang mendekati 1 menunjukkan bahwa model mampu merepresentasikan pola data dengan baik, sedangkan nilai negatif mengindikasikan bahwa model memberikan hasil yang lebih buruk dibandingkan pendekatan sederhana menggunakan nilai rata-rata.

MAE dihitung sebagai rata-rata selisih absolut antara nilai observasi dan nilai hasil peramalan, sehingga metrik ini memberikan ukuran kesalahan yang mudah diinterpretasikan dalam satuan PM2.5. MAE juga relatif lebih *robust* terhadap keberadaan pencilan karena tidak memberikan penalti berlebih pada

kesalahan ekstrem. Sementara itu, RMSE diperoleh dari akar kuadrat rata-rata selisih kuadrat antara nilai aktual dan nilai peramalan. Berbeda dengan MAE, RMSE lebih sensitif terhadap kesalahan besar karena memberikan bobot lebih tinggi pada deviasi yang ekstrem, sehingga metrik ini berguna untuk menilai stabilitas model dalam menghadapi fluktuasi tajam pada kualitas udara.

Seluruh metrik dihitung pada data uji di setiap lipatan validasi berbasis waktu tanpa pengacakan. Hasil evaluasi disajikan dalam bentuk rerata untuk masing-masing wilayah, dan bila diperlukan dirangkum secara lintas wilayah. Sebagai konteks pembandingan, *baseline persistence* yang menggunakan nilai terakhir turut dihitung pada lipatan yang sama, meskipun fokus analisis utama tetap diarahkan pada performa model SVR.

### 3.4 Perhitungan Manual

Pada penelitian ini, perhitungan manual disusun untuk menjelaskan secara transparan tahapan utama yang dilakukan sistem dalam membangun model peramalan PM2.5. Uraian pada subbab ini bersifat ilustratif untuk memudahkan pemahaman terhadap proses pembentukan fitur, normalisasi, seleksi fitur, hingga pemodelan. Sementara itu, implementasi aktual pada sistem dilakukan secara otomatis melalui *pipeline* yang telah dibangun pada Laravel dan Flask menggunakan data historis nyata dari setiap wilayah pengamatan.

Proses pada sistem aktual dimulai dari deret waktu PM2.5 per wilayah yang tersimpan di basis data MySQL. Data kemudian diurutkan secara kronologis berdasarkan waktu pengamatan lokal (*observed\_at*), diproses melalui tahapan praproses, rekayasa fitur, transformasi, seleksi fitur, pelatihan model, dan inferensi

secara otomatis. Dengan demikian, contoh perhitungan manual pada subbab ini tidak merepresentasikan keseluruhan volume data penelitian, melainkan digunakan sebagai ilustrasi sederhana untuk menunjukkan logika kerja sistem.

Target peramalan didefinisikan sebagai nilai PM2.5 pada hari berikutnya dengan rentang 0 hingga 500. Fitur yang digunakan mencakup *lag* pada variabel target maupun variabel independen meteorologi, seperti *pm25\_lag1*, *t\_lag1*, dan *h\_lag1*, serta fitur musiman harian berupa pengodean hari dalam pekan. Seluruh transformasi data ditempatkan di dalam alur validasi untuk mencegah kebocoran informasi, meliputi penskalaan fitur, serta transformasi target yang bersifat monotonik agar stabil pada nilai nol maupun negatif.

Pemilihan konfigurasi model dilakukan melalui pencarian parameter secara terstruktur pada ruang kandidat dengan pembagian data berbasis waktu tanpa pengacakan. Kinerja model dievaluasi menggunakan *walk-forward validation*. Setelah penalaan selesai, model dilatih ulang menggunakan seluruh data historis dengan skema bobot waktu yang sama. Hasil peramalan H+1 kemudian dikembalikan ke skala asli PM2.5 melalui *inverse transform*, lalu dihitung menjadi IAQI dan diklasifikasikan ke dalam kategori kualitas udara berdasarkan standar AQI US. Selain itu, nilai PM2.5 yang sama juga digunakan untuk menentukan kategori ISPU sesuai *breakpoint* yang berlaku.

### 3.4.1 Data Awal

Data yang digunakan dalam penelitian ini merupakan hasil pengukuran parameter kualitas udara dan variabel meteorologi di wilayah pengamatan. Variabel yang tersedia meliputi PM2.5 (*pm25*), suhu udara (*t*), kelembapan (*h*), tekanan

udara (p), kecepatan angin (w), dan titik embun (dew). Seluruh data telah lengkap dan tersimpan dalam basis data MySQL dengan format waktu pengamatan `observed_at`, diurutkan secara kronologis berdasarkan waktu lokal.

Tabel 3.2 Contoh Data *Dummy*

Tanggal	pm25	t	h	p	w	dew
2025-10-24	154	25.5	92	1009.1	3	24.1
2025-10-25	159	25.9	95.4	1009.2	2.5	25.1
2025-10-26	65	26	100	1009	2	26
2025-10-27	69	26.7	95.8	1010	2.1	25.1
2025-10-28	109	26	94	1010	1.5	25

Data pada Tabel 3.3 digunakan sebagai ilustrasi sederhana untuk menjelaskan tahapan rekayasa fitur, yang meliputi pembentukan *lag*, *rolling mean*, *first difference*, EWMA, dan fitur musiman. Adapun pada implementasi aktual, seluruh tahapan tersebut dijalankan secara otomatis pada dataset historis nyata di dalam sistem.

### 3.4.2 Rekayasa Fitur

Pada tahap awal, dilakukan rekayasa fitur yang akan digunakan dalam pemodelan. Proses ini dimulai dari data deret waktu hasil pengukuran parameter kualitas udara dan cuaca yang telah lengkap, meliputi variabel PM2.5 (pm25), suhu (t), kelembapan (h), tekanan udara (p), kecepatan angin (w), dan suhu titik embun (dew). Data tersimpan dalam basis data MySQL dengan format waktu pengamatan `observed_at` (misalnya "2025-10-28"), dan diurutkan secara kronologis berdasarkan waktu lokal.

Fitur-fitur utama yang dibentuk meliputi *lag*, *rolling mean*, *first difference*, *Exponential Weighted Moving Average* (EWMA) dan fitur musiman harian. Fitur

*lag* digunakan untuk menangkap pola temporal jangka pendek dengan mengambil nilai pengamatan pada waktu sebelumnya ( $t-1$ ). Misalnya, *pm25\_lag1* menunjukkan nilai PM2.5 pada waktu sebelumnya, *t\_lag1* adalah suhu pada data observasi sebelumnya, dan demikian pula untuk kelembapan (*h\_lag1*), tekanan udara (*p\_lag1*), kecepatan angin (*w\_lag1*), serta titik embun (*dew\_lag1*). Selain itu, ditambahkan fitur *rolling mean* untuk menangkap tren jangka pendek atau fluktuasi harian dalam variabel. Misalnya, *pm25\_rolling3* adalah rata-rata PM2.5 dari 3 data observasi terakhir. Rumus *rolling window* sederhana untuk  $j$  hari terakhir adalah:

$$x_t^{\text{rolling-}j} = \frac{1}{j} \sum_{k=0}^{j-1} x_{t-k} \quad (3.1)$$

Keterangan:

$x_t^{\text{rolling-}j}$  : nilai rata-rata variabel  $x$  pada tanggal  $t$  berdasarkan  $j$  data terakhir  
 $x_{t-k}$  : nilai variabel  $x$  pada data ke- $t-k$   
 $j$  : jumlah data untuk *rolling* (misal 3 atau 7)

Selain itu, ditambahkan fitur musiman berupa pengodean siklik hari dalam minggu menggunakan fungsi sinus dan kosinus untuk mempertahankan kontinuitas antar hari, sebagaimana dirumuskan berikut:

$$\text{dow\_sin}_t = \sin\left(2\pi \cdot \frac{\text{dow}_t}{7}\right) \quad (3.2)$$

$$\text{dow\_cos}_t = \cos\left(2\pi \cdot \frac{\text{dow}_t}{7}\right) \quad (3.3)$$

Keterangan:

$\text{dow\_sin}_t$  : komponen sinus untuk pengodean siklik hari-dalam-minggu pada tanggal  $t$   
 $\text{dow\_cos}_t$  : komponen kosinus untuk pengodean siklik hari-dalam-minggu pada tanggal  $t$   
 $2\pi$  : faktor periodik yang memastikan periode tepat 7 hari  
 $\text{dow}_t$  : indeks hari dalam minggu pada tanggal  $t$  (Senin = 1, ..., Minggu = 7)

Untuk menangkap tingkat perubahan (*rate of change*) PM2.5 dari satu hari ke hari berikutnya, dibentuk fitur *first difference* yang dihitung sebagai selisih nilai PM2.5 antara waktu  $t$  dan  $t - 1$ :

$$pm25\_diff1_t = pm25_t - pm25_{t-1} \quad (3.4)$$

Keterangan:

$pm25\_diff1_t$  : selisih nilai PM2.5 antara hari  $t$  dan  $t - 1$  (tingkat perubahan harian)

$pm25_t$  : nilai PM2.5 pada waktu  $t$

$pm25_{t-1}$  : nilai PM2.5 pada waktu  $t - 1$

Fitur ini membantu model menangkap dinamika perubahan mendadak dalam kualitas udara, seperti lonjakan atau penurunan PM2.5 yang signifikan. Selain *first difference*, dibentuk juga fitur EWMA yang memberikan bobot lebih besar pada pengamatan terbaru dan bobot yang menurun secara eksponensial untuk pengamatan yang lebih lama. EWMA dihitung baik untuk variabel target PM2.5 maupun variabel independen meteorologi menggunakan rumus rekursif:

$$EWMA_t = \alpha \cdot x_t + (1 - \alpha) \cdot EWMA_{t-1} \quad (3.5)$$

Keterangan:

$EWMA_t$  : nilai *exponentially weighted moving average* pada waktu  $t$

$\alpha$  : parameter smoothing (biasanya  $\alpha = \frac{2}{W+1}$ , di mana  $W$  adalah jendela)

$x_t$  : nilai observasi pada waktu  $t$

$EWMA_{t-1}$  : nilai EWMA pada waktu sebelumnya

Setelah fitur-fitur ini dihitung, data tersebut digunakan untuk membangun model peramalan PM2.5 menggunakan algoritma *Support Vector Regression* (SVR). Data yang telah diproses ini akan dimasukkan ke dalam model untuk meramalkan nilai PM2.5 pada hari berikutnya ( $H+1$ ) berdasarkan pola yang ditemukan dalam data historis dan fitur-fitur yang terbentuk.



Representasi pada Lampiran 1 menunjukkan bahwa setiap tanggal (T) memiliki vektor fitur yang memadukan dependensi temporal jangka pendek melalui komponen *lag* ( $pm25\_lag1, t\_lag1, h\_lag1, p\_lag1, w\_lag1, dew\_lag1$ ) dan pola musiman mingguan melalui pasangan ( $dow\_sin_t, dow\_cos_t$ ) sebagaimana didefinisikan pada Persamaan (3.2) - (3.3). Dengan demikian, setiap baris pada tanggal dipetakan ke vektor  $X_t$  yang digunakan untuk meramalkan target  $Y_{t+1}$  (PM2.5 pada tanggal berikutnya), yaitu  $Y_{t+1} = f(x_t)$ . Formulasi ini memastikan bahwa informasi yang relevan baik dinamika periodisitas mingguan terikut secara konsisten ke dalam proses pemodelan. Pada bagian berikutnya akan diuraikan tahapan prapemodelan yang meliputi penskalaan fitur serta transformasi target yang bersifat monotonik sebelum estimasi dilakukan menggunakan algoritma *Support Vector Regression* (SVR).

### 3.4.3 Normalisasi Data dan Seleksi Fitur

Sebelum data digunakan dalam proses pemodelan, seluruh variabel independen terlebih dahulu dinormalisasi untuk mengurangi efek perbedaan skala antar fitur. Normalisasi diperlukan agar variabel dengan rentang nilai besar (misalnya tekanan udara) tidak mendominasi variabel lain seperti suhu atau kecepatan angin dalam proses pembelajaran model. Proses normalisasi dilakukan dengan rumus sebagai berikut:

$$Z = \frac{x - \mu}{\sigma} \quad (3.6)$$

Keterangan:

- $Z$  : data yang sudah dinormalisasikan
- $x$  : nilai asli fitur
- $\mu$  : rata-rata dari semua nilai fitur tersebut
- $\sigma$  : Standar deviasi dari semua nilai fitur tersebut

Proses normalisasi menggunakan *StandardScaler* dilakukan untuk menyamakan skala antar fitur agar tidak terjadi dominasi variabel tertentu. Pada Lampiran 2 menampilkan hasil normalisasi pada data *dummy* sebagai ilustrasi proses, sedangkan normalisasi pada sistem aktual dilakukan secara otomatis pada setiap lipatan validasi. Setelah seluruh fitur berada pada skala yang seragam, tahap berikutnya adalah seleksi fitur untuk memilih subset variabel yang memiliki hubungan paling signifikan secara statistik terhadap target peramalan, yaitu PM2.5 pada hari berikutnya. Seleksi fitur dilakukan menggunakan metode *SelectKBest* dengan fungsi *scoring f\_regression*, yang menghitung *F-statistic* berdasarkan uji ANOVA (*Analysis of Variance*) untuk regresi. *F-statistic* mengukur signifikansi hubungan linear antara setiap fitur dengan target, dengan memperhitungkan ukuran sampel dan variabilitas data. Rumus *F-statistic* untuk seleksi fitur dituliskan sebagai berikut:

$$F_j = \frac{r_j^2}{1 - r_j^2} \times (n - 2) \quad (3.7)$$

Keterangan:

- $F_j$  : F-statistic untuk fitur ke-j  
 $r_j$  : koefisien korelasi Pearson antara fitur  $x_j$  dan target  $y$   
 $n$  : jumlah sampel pengamatan

Hasil perhitungan korelasi digunakan untuk menentukan fitur yang relevan, selanjutnya dilakukan seleksi fitur dengan memilih  $k$  fitur berdasarkan nilai *F-statistic* tertinggi, sementara fitur dengan nilai rendah dieliminasi karena kontribusinya terhadap variabel target terbatas.

Tabel 3.3 Hasil Seleksi Fitur terhadap Target (Data *Dummy*)

Fitur	Nilai Korelasi ( $r_i$ )	$F_i$	Status
pm25_ewma'	0.91	15.51	Terpilih

pm25_diff1'	-0.90	12.15	Terpilih
pm25_lag1'	0.81	8.77	Terpilih
pm25_rmean3'	0.79	7.04	Terpilih
w'	0.42	0.95	Terpilih
h'	-0.32	0.41	Terpilih
dew'	-0.15	0.08	Terpilih
t'	0.13	0.06	Terpilih
p'	-0.14	0.07	Tidak Terpilih
t_lag1'	0.05	0.01	Tidak Terpilih
h_lag1'	-0.02	0.00	Tidak Terpilih
p_lag1'	-0.10	0.03	Tidak Terpilih
w_lag1'	0.30	0.33	Tidak Terpilih
dew_lag1'	-0.07	0.02	Tidak Terpilih
t_rmean3'	-0.05	0.01	Tidak Terpilih
h_rmean3'	0.08	0.02	Tidak Terpilih
p_rmean3'	-0.09	0.03	Tidak Terpilih
w_rmean3'	0.10	0.03	Tidak Terpilih
dew_rmean3'	-0.06	0.01	Tidak Terpilih
t_ewma'	-0.03	0.00	Tidak Terpilih
h_ewma'	0.07	0.02	Tidak Terpilih
p_ewma'	-0.02	0.00	Tidak Terpilih
w_ewma'	0.05	0.01	Tidak Terpilih
dew_ewma'	-0.01	0.00	Tidak Terpilih
dow_sin'	0.10	0.03	Tidak Terpilih
dow_cos'	-0.01	0.00	Tidak Terpilih

Tabel 3.4 disajikan sebagai ilustrasi hasil seleksi fitur pada data *dummy*, sedangkan seleksi fitur pada implementasi aktual dilakukan secara otomatis di dalam *pipeline* model. Dari hasil seleksi pada Tabel 3.4, terlihat bahwa fitur dengan korelasi tinggi seperti *pm25\_ewma*, *pm25\_diff1*, *pm25\_lag1*, *pm25\_rmean3*, *w*, *h*, *dew*, *t*, memiliki hubungan yang kuat terhadap target dan tetap digunakan dalam pemodelan. Sementara itu, fitur dengan kontribusi lemah atau redundan dieliminasi untuk mencegah *overfitting* dan meningkatkan efisiensi komputasi. Dengan demikian, data hasil seleksi fitur inilah yang digunakan dalam proses pemodelan pada tahap berikutnya.

### 3.4.4 Pemodelan dan Peramalan

Sistem menerapkan prosedur pemodelan *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF) untuk meramalkan nilai PM2.5 pada hari berikutnya (H+1). Data yang digunakan telah melalui tahapan praproses, termasuk rekayasa fitur, serta normalisasi menggunakan *StandardScaler* sebagaimana dijelaskan pada subbab sebelumnya. Setelah proses normalisasi dilakukan, seluruh fitur akan diseleksi menggunakan *SelectKBest* dengan fungsi skor *f\_regression*. Target peramalan dalam penelitian ini adalah nilai PM2.5 pada horizon satu hari ke depan. Model SVR dengan kernel RBF dipelajari dari data latih  $\{(x_i, y_i)\}_{i=1}^n$ , dengan peramalan untuk input baru  $x$  dihitung sebagai:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad (3.8)$$

Dengan kernel RBF:

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \quad (3.9)$$

Keterangan:

$\alpha_i$	: bobot Lagrange yang diperoleh dari optimisasi SVR,
$b$	: bias/intercept,
$\gamma$	: parameter kernel yang mengontrol lebar fungsi RBF,
$\ x_i - x\ ^2$	: jarak Euclidean antara vektor fitur latih dan input baru

Penentuan hiperparameter utama SVR, yaitu  $C$ ,  $\epsilon$ , dan  $\gamma$ , dilakukan melalui pencarian parameter terstruktur (*grid search*) dengan pembagian data berbasis waktu. Validasi dilakukan menggunakan skema *walk-forward validation*, di mana model dilatih pada data historis dan diuji pada data yang lebih baru, sehingga proses evaluasi tetap mencerminkan kondisi peramalan nyata.

Setelah model terbaik diperoleh, sistem melakukan pelatihan akhir menggunakan seluruh data historis yang tersedia pada masing-masing wilayah. *Output* model berupa peramalan konsentrasi PM2.5 kemudian dikonversi menjadi indeks kualitas udara menggunakan rumus interpolasi linier US EPA maupun standar ISPU nasional. Dengan demikian, hasil peramalan tidak hanya berupa angka konsentrasi, tetapi juga dapat langsung diinterpretasikan dalam kategori mutu udara yang berlaku.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Gambaran Umum Penelitian**

Bab ini menyajikan hasil implementasi sistem serta evaluasi model peramalan kualitas udara yang dikembangkan dalam penelitian ini. Fokus utama penelitian adalah melakukan peramalan konsentrasi partikulat halus PM2.5 untuk satu hari ke depan (H+1) menggunakan metode *Support Vector Regression* (SVR). Estimasi harian tersebut diperoleh dengan strategi *recursive multi-step forecasting* sebanyak 24 langkah yang kemudian dirata-ratakan. Nilai PM2.5 hasil peramalan kemudian digunakan untuk menghitung indeks kualitas udara IAQI yang selanjutnya dikategorikan berdasarkan standar US EPA dan ISPU Indonesia. Selain membahas performa model peramalan, bab ini juga menguraikan bagaimana hasil peramalan tersebut diintegrasikan dan ditampilkan melalui sistem informasi berbasis web. Dengan demikian, Bab ini tidak hanya menitikberatkan pada aspek pemodelan, tetapi juga pada hasil implementasi sistem secara keseluruhan.

Data yang digunakan dalam penelitian ini merupakan data historis kualitas udara yang diperoleh melalui *Application Programming Interface* (API) *World Air Quality Index* (WAQI). Karakteristik data yang diperoleh bersifat tidak seragam, karena pada beberapa hari hanya tersedia satu pengamatan, sedangkan pada hari lainnya terdapat beberapa pengamatan dalam satu hari. Rentang waktu pengumpulan data dimulai dari tanggal 27 Juni 2025 hingga 22 Januari 2026, yang kemudian digunakan sebagai dasar untuk menghasilkan peramalan pada tanggal 23

Januari 2026. Keterbatasan jumlah dan resolusi temporal data ini menjadi salah satu faktor penting yang memengaruhi hasil peramalan dan akan dibahas lebih lanjut pada bagian pembahasan.

Selain evaluasi model peramalan, Bab ini juga memaparkan hasil implementasi sistem informasi pemantauan kualitas udara berbasis web. Sistem ini dibangun menggunakan *framework* Laravel sebagai *backend* dan penyaji antarmuka pengguna, serta Flask sebagai layanan peramalan yang menangani proses *machine learning*. Sistem mampu menampilkan kondisi kualitas udara terkini berdasarkan data WAQI, hasil peramalan PM2.5 satu hari ke depan, serta konversinya ke dalam kategori indeks kualitas udara. Penyajian informasi dilakukan melalui tabel dan peta interaktif untuk memudahkan pengguna dalam memahami kondisi kualitas udara di berbagai wilayah. Dengan adanya sistem ini, hasil penelitian tidak hanya bersifat analitis, tetapi juga diwujudkan dalam bentuk aplikasi yang dapat digunakan sebagai sarana pemantauan dan informasi awal kualitas udara.

#### **4.2 Karakteristik Data Penelitian**

Subbab ini menjelaskan karakteristik data yang digunakan dalam proses peramalan kualitas udara, meliputi sumber data, rentang waktu pengambilan data, serta variasi jumlah data pada setiap wilayah penelitian. Pemahaman terhadap karakteristik data menjadi hal yang penting karena kualitas dan struktur data sangat berpengaruh terhadap kinerja model peramalan yang dibangun. Data yang digunakan dalam penelitian ini bersifat deret waktu (*time series*), sehingga pola temporal yang dapat dipelajari oleh model memiliki keterbatasan tertentu. Selain

itu, ketersediaan data yang tidak seragam antar wilayah menyebabkan jumlah data historis yang digunakan pada setiap wilayah berbeda-beda. Kondisi ini menjadi salah satu faktor yang memengaruhi hasil evaluasi model dan akan dibahas lebih lanjut pada bagian pembahasan.

Karakteristik data dalam penelitian ini juga mencerminkan kondisi nyata dari sistem pengambilan data kualitas udara berbasis API, di mana peneliti tidak memiliki kendali penuh terhadap frekuensi dan kelengkapan data. Data yang diperoleh dari WAQI API merepresentasikan hasil pengamatan dari stasiun pemantauan kualitas udara yang tersebar di berbagai wilayah. Setiap wilayah memiliki tingkat kelengkapan data yang berbeda, bergantung pada stabilitas stasiun pemantauan dan ketersediaan sensor. Oleh karena itu, karakteristik data pada penelitian ini bersifat heterogen dan tidak sepenuhnya ideal untuk pemodelan deret waktu. Kondisi tersebut menjadi konteks penting dalam mengevaluasi performa model peramalan yang dihasilkan.

#### **4.2.1 Sumber dan Rentang Data**

Data kualitas udara yang digunakan dalam penelitian ini diperoleh dari *World Air Quality Index (WAQI)* melalui layanan *Application Programming Interface (API)*. Data diambil untuk beberapa wilayah di Indonesia yang memiliki ketersediaan data PM<sub>2.5</sub> dan parameter meteorologi pendukung. Pengambilan data dilakukan secara periodik selama periode penelitian, dengan karakteristik data dalam bentuk pengamatan deret waktu dengan frekuensi yang tidak selalu seragam. Pada beberapa hari hanya tersedia satu pengamatan, sedangkan pada hari lainnya terdapat lebih dari satu pengamatan dalam sehari. Data yang digunakan dalam



penelitian ini mencakup periode 27 Juni 2025 hingga 22 Januari 2026. Berdasarkan data historis tersebut, model peramalan digunakan untuk menghasilkan peramalan kualitas udara satu hari ke depan (H+1) pada tanggal 23 Januari 2026.

Tabel 4.1 Ringkasan Data Penelitian per Wilayah

No	Wilayah/Kota	Tanggal Awal Pengambilan	Tanggal Akhir Pengambilan	Jumlah Data Tersedia
1	Bengkulu	2025-06-27 13:00:00	2026-01-22 23:00:00	1902
2	Pekanbaru	2025-06-27 13:00:00	2026-01-22 23:00:00	1882
3	Sleman	2025-06-27 13:00:00	2026-01-22 23:00:00	1738
4	Malang	2025-06-27 13:00:00	2026-01-22 22:00:00	1857
5	Medan	2025-06-27 13:00:00	2026-01-22 23:00:00	1978
6	Kemayoran	2025-06-27 13:00:00	2026-01-22 23:00:00	1889
7	Banjarbaru	2025-06-27 13:00:00	2026-01-22 23:00:00	1954
8	Samarinda	2025-06-27 13:00:00	2026-01-22 23:00:00	1914
9	Jambi	2025-06-27 13:00:00	2026-01-22 23:00:00	1956
10	Semarang	2025-06-27 13:00:00	2026-01-22 23:00:00	1739
11	Sorong	2025-07-07 17:00:00	2026-01-22 23:00:00	1363
12	Kototabang	2025-06-27 13:00:00	2026-01-22 22:00:00	1542
13	Talang Betutu	2025-06-27 13:00:00	2026-01-22 23:00:00	1972
14	Pangkalanbun	2025-06-27 13:00:00	2026-01-22 23:00:00	1916
15	Maros	2025-06-27 13:00:00	2026-01-22 23:00:00	1630

Tabel 4.1 menunjukkan ringkasan data historis kualitas udara yang digunakan dalam penelitian ini pada masing-masing wilayah. Terlihat bahwa rentang waktu pengambilan data dan jumlah data yang tersedia berbeda-beda antarwilayah. Perbedaan ini disebabkan oleh variasi ketersediaan data pada stasiun pemantauan kualitas udara yang terhubung dengan WAQI API. Perbedaan jumlah data ini dapat memberikan konteks yang berbeda dalam proses pelatihan dan evaluasi model peramalan, meskipun kinerja model juga dipengaruhi oleh faktor lain seperti karakteristik distribusi data, kualitas data, dan tahap prapemrosesan yang dilakukan.

#### 4.2.2 Variabel yang Digunakan

Pemilihan variabel dalam penelitian ini dilakukan berdasarkan tingkat ketersediaan dan konsistensi data yang diperoleh dari *World Air Quality Index* (WAQI) API pada seluruh wilayah penelitian. Hasil eksplorasi data awal menunjukkan bahwa tidak semua parameter kualitas udara dan meteorologi tersedia secara lengkap untuk setiap wilayah dan setiap waktu pengamatan. Oleh karena itu, penelitian ini memfokuskan penggunaan variabel yang secara konsisten terisi pada seluruh wilayah selama periode penelitian. Pendekatan ini bertujuan untuk menjaga keseragaman fitur input antar wilayah serta menghindari pengurangan jumlah data efektif. Dengan demikian, hanya variabel yang memenuhi kriteria kelengkapan dan kontinuitas data yang digunakan dalam proses peramalan.

Variabel yang digunakan dalam penelitian ini meliputi konsentrasi PM2.5 sebagai variabel target, serta beberapa variabel meteorologi sebagai fitur pendukung. Ringkasan variabel yang digunakan, termasuk satuan dan peran masing-masing variabel dalam model peramalan, disajikan pada Tabel 4.2.

Tabel 4.2 Variabel Penelitian yang Digunakan dalam Model Peramalan

No	Variabel	Simbol	Jenis Variabel	Keterangan
1	Konsentrasi PM2.5	PM2.5	Target	Parameter utama kualitas udara
2	Suhu	t	Fitur	Kondisi termal lingkungan
3	Kelembapan Udara	h	Fitur	Kadar uap air di udara
4	Tekanan Udara	p	Fitur	Tekanan atmosfer
5	Kecepatan Angin	w	Fitur	Faktor dispersi polutan
6	Titik Embun	dew	Fitur	Indikator kelembapan absolut

Berdasarkan Tabel 4.2, PM2.5 digunakan sebagai variabel target karena merupakan parameter utama dalam penentuan indeks kualitas udara dan memiliki dampak signifikan terhadap kesehatan manusia. Variabel meteorologi, seperti suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun, digunakan sebagai fitur pendukung karena berperan dalam memengaruhi proses pembentukan, akumulasi, dan dispersi partikel PM2.5 di atmosfer. Seluruh variabel tersebut diperoleh langsung dari WAQI API dan digunakan tanpa interpolasi tambahan. Variabel curah hujan tidak disertakan dalam penelitian ini karena ketersediaannya tidak konsisten pada seluruh wilayah penelitian. Pengecualian variabel tersebut dilakukan untuk menjaga stabilitas proses pemodelan.

#### **4.2.3 Resolusi Temporal Data**

Data yang digunakan dalam penelitian ini memiliki resolusi temporal yang tidak sepenuhnya seragam, karena jumlah pengamatan yang tersedia pada setiap hari dapat berbeda-beda. Pada beberapa periode, data kualitas udara hanya tercatat satu kali dalam sehari, sedangkan pada hari lainnya dapat terdapat lebih dari satu pengamatan dalam satu hari. Pola ini merupakan karakteristik bawaan dari data yang disediakan oleh *World Air Quality Index (WAQI) API*, yang tidak selalu memberikan frekuensi pengamatan yang konsisten untuk seluruh wilayah penelitian. Setiap nilai pengamatan merepresentasikan kondisi kualitas udara pada waktu tertentu, sehingga dataset yang digunakan mencerminkan tren PM2.5 berdasarkan data historis yang tersedia. Oleh karena itu, data penelitian ini lebih tepat digolongkan sebagai deret waktu dengan frekuensi pengamatan yang bervariasi (*irregular time series*).

Ketidakteraturan resolusi temporal ini berdampak pada kemampuan model dalam menangkap fluktuasi PM<sub>2.5</sub> yang dipengaruhi oleh aktivitas manusia dan perubahan kondisi meteorologi dalam satu hari. Aktivitas lalu lintas, industri, serta perubahan cuaca yang terjadi pada waktu tertentu tidak selalu tercermin secara merata dalam data. Akibatnya, pola lonjakan atau penurunan PM<sub>2.5</sub> dalam skala jam tidak dapat dimodelkan secara optimal. Hal ini menjadi salah satu faktor yang memengaruhi rendahnya nilai koefisien determinasi ( $R^2$ ) pada hasil evaluasi model peramalan. Meskipun demikian, penggunaan seluruh data historis yang tersedia masih relevan untuk peramalan jangka pendek ( $H+1$ ) yang bertujuan memberikan gambaran umum kondisi kualitas udara pada hari berikutnya.

Meskipun frekuensi data tidak selalu konsisten, dataset ini tetap memungkinkan penerapan pendekatan deret waktu dengan memanfaatkan pola historis antar pengamatan. Model peramalan dalam penelitian ini difokuskan untuk menangkap kecenderungan perubahan PM<sub>2.5</sub> dari waktu ke waktu berdasarkan data yang tersedia, bukan fluktuasi *intraday* secara penuh. Pendekatan ini sejalan dengan tujuan sistem informasi yang dikembangkan, yaitu memberikan informasi prakiraan kualitas udara satu hari ke depan sebagai referensi awal bagi masyarakat dan instansi terkait. Dengan demikian, keterbatasan resolusi temporal tidak sepenuhnya menghilangkan nilai guna sistem, tetapi membatasi tingkat presisi peramalan. Keterbatasan ini menjadi dasar pertimbangan dalam pemberian saran pengembangan penelitian pada bab selanjutnya.

### 4.3 Pra-pemrosesan dan Rekayasa Fitur

Tahapan pra-pemrosesan dan rekayasa fitur merupakan langkah penting dalam penelitian ini untuk memastikan data yang digunakan layak dan optimal bagi proses peramalan menggunakan *Support Vector Regression (SVR)*. Mengingat data kualitas udara yang digunakan memiliki frekuensi pengamatan yang tidak selalu seragam serta jumlah data yang relatif terbatas pada beberapa wilayah, maka pengolahan data dilakukan secara hati-hati agar tidak mengurangi jumlah sampel yang tersedia secara signifikan. Fokus utama pada tahap ini adalah menjaga konsistensi deret waktu, mengurangi *noise*, serta mengekstraksi informasi temporal yang relevan dari data historis. Dengan pra-pemrosesan dan rekayasa fitur yang tepat, diharapkan model mampu mempelajari pola perubahan PM2.5 antar waktu pengamatan secara lebih efektif.

Selain itu, tahapan ini juga bertujuan untuk mengatasi karakteristik data kualitas udara yang sering kali mengandung perbedaan skala antar variabel, serta fluktuasi yang tidak teratur. Proses pra-pemrosesan dilakukan sebelum data digunakan dalam pembentukan dataset *supervised learning*. Selanjutnya, rekayasa fitur diterapkan untuk memperkaya representasi data tanpa menambah ketergantungan pada variabel eksternal yang tidak tersedia secara konsisten. Pendekatan ini disesuaikan dengan keterbatasan resolusi temporal dan jumlah data yang tersedia pada masing-masing wilayah penelitian.

#### 4.3.1 Pra-pemrosesan Data

Pra-pemrosesan data pada penelitian ini dilakukan untuk memastikan bahwa data kualitas udara yang digunakan berada dalam bentuk yang konsisten dan

siap digunakan dalam pemodelan deret waktu. Tahapan pertama yang dilakukan adalah pengurutan data berdasarkan waktu pengamatan (*time-series ordering*) untuk menjaga urutan kronologis data. Pengurutan ini sangat penting karena metode peramalan yang digunakan bergantung pada hubungan antar waktu, sehingga kesalahan urutan data dapat menyebabkan bias pada hasil peramalan. Setiap data pengamatan diurutkan berdasarkan tanggal observasi yang diperoleh dari WAQI API.

Tahap berikutnya adalah penanganan data duplikat dan data tidak valid. Data dengan waktu pengamatan yang sama disimpan satu kali untuk menghindari redundansi. Selanjutnya, dilakukan penghapusan baris data yang tidak memiliki nilai PM2.5, karena variabel tersebut merupakan target utama dalam proses peramalan. Pendekatan ini dipilih agar model tidak dilatih menggunakan target yang tidak lengkap atau tidak valid. Dengan demikian, hanya data yang memiliki nilai PM2.5 yang digunakan dalam proses selanjutnya.

Variabel yang tidak tersedia secara konsisten pada seluruh wilayah, seperti curah hujan, tidak disertakan sejak awal dalam dataset. Pendekatan ini dipilih untuk menghindari penghapusan baris data secara berlebihan yang dapat mengurangi jumlah sampel pelatihan. Setelah seluruh tahapan pra-pemrosesan dilakukan, data yang telah dibersihkan membentuk deret waktu yang siap digunakan dalam proses rekayasa fitur dan pembentukan dataset *supervised learning*. Jumlah data akhir yang digunakan pada setiap wilayah bervariasi, tergantung pada ketersediaan data PM2.5 yang valid. Data hasil pra-pemrosesan ini menjadi dasar utama dalam tahap

rekayasa fitur dan pelatihan model peramalan kualitas udara pada subbab berikutnya.

#### 4.3.2 Rekayasa Fitur

Rekayasa fitur dilakukan untuk mengekstraksi informasi temporal yang relevan dari data kualitas udara yang tersedia. Mengingat data memiliki frekuensi pengamatan yang tidak seragam, rekayasa fitur difokuskan pada hubungan antar pengamatan sebelumnya (*lag*) dan kecenderungan perubahan jangka pendek. Pendekatan ini bertujuan untuk membantu model memahami kecenderungan perubahan konsentrasi PM<sub>2.5</sub> berdasarkan nilai historisnya. Fitur-fitur yang dibentuk dirancang agar mampu merepresentasikan pola tren, arah perubahan, serta kestabilan PM<sub>2.5</sub> dari waktu ke waktu. Dengan rekayasa fitur yang tepat, keterbatasan resolusi temporal dapat diminimalkan tanpa menambah ketergantungan pada data eksternal.

Fitur utama yang digunakan meliputi *lag* PM<sub>2.5</sub> berbasis langkah waktu ( $lag-1$  = satu data observasi sebelumnya), *rolling mean*, *first difference* antar observasi berturutan, serta *Exponential Weighted Moving Average* (EWMA) yang memberi bobot lebih besar pada data terbaru. Kombinasi fitur *lag*, *rolling mean*, *first difference*, dan EWMA memungkinkan model menangkap dinamika jangka pendek dan tren umum PM<sub>2.5</sub> pada resolusi waktu data.

Selain fitur berbasis PM<sub>2.5</sub>, penelitian ini juga memasukkan fitur meteorologi yang telah direkayasa secara sederhana. Variabel suhu, kelembapan, tekanan udara, kecepatan angin, dan titik embun digunakan dalam bentuk nilai *lag* dan EWMA untuk menjaga konsistensi dengan fitur PM<sub>2.5</sub>. Rekayasa fitur

meteorologi dilakukan secara terbatas untuk menghindari peningkatan jumlah fitur yang berlebihan, mengingat jumlah data pelatihan yang relatif kecil. Pendekatan ini bertujuan untuk menyeimbangkan antara kelengkapan informasi dan risiko *overfitting*. Secara keseluruhan, proses rekayasa fitur menghasilkan sejumlah fitur numerik yang mampu merepresentasikan dinamika kualitas udara secara lebih informatif dan stabil. Fitur-fitur hasil rekayasa ini kemudian digunakan sebagai input dalam pembentukan dataset *supervised learning* untuk pelatihan model *Support Vector Regression*.

#### 4.3.3 Seleksi Fitur

Proses seleksi fitur dilakukan untuk mengurangi kompleksitas model serta meminimalkan risiko *overfitting* akibat jumlah fitur hasil rekayasa yang relatif besar dibandingkan dengan jumlah data historis yang tersedia pada masing-masing wilayah. Setelah tahap rekayasa fitur, diperoleh sebanyak 26 fitur numerik yang mencakup fitur berbasis histori PM2.5 serta variabel meteorologi. Mengingat jumlah data observasi pada setiap wilayah tidak terlalu besar, penggunaan seluruh fitur tersebut berpotensi menurunkan kemampuan generalisasi model. Oleh karena itu, seleksi fitur dilakukan untuk mempertahankan hanya fitur-fitur yang paling relevan terhadap target peramalan PM2.5 pada periode observasi berikutnya. Tahapan ini menjadi penting agar model dapat bekerja secara lebih stabil, efisien, dan tidak terlalu sensitif terhadap *noise* pada data.

Metode seleksi fitur yang digunakan adalah *SelectKBest* dengan fungsi skor  $f\_regression$ , yang mengukur kekuatan hubungan linear antara masing-masing fitur dengan variabel target PM2.5. Berdasarkan hasil Eksperimen 3, sistem



menggunakan konfigurasi  $k = 8$ , yang berarti pada setiap wilayah dipilih delapan fitur terbaik dari total 26 fitur hasil rekayasa. Pendekatan ini terbukti mampu meningkatkan performa model pada sebagian besar wilayah dibandingkan dengan penggunaan seluruh fitur. Proses seleksi fitur ini diintegrasikan langsung ke dalam *pipeline* model *Support Vector Regression* sehingga dilakukan secara otomatis pada setiap proses pelatihan.

Meskipun jumlah fitur yang dipilih sama (delapan fitur), identitas fitur yang terpilih dapat berbeda antar wilayah bergantung pada karakteristik data masing-masing lokasi. Tabel 4.3 menyajikan ringkasan hasil seleksi fitur pada 15 wilayah penelitian. Proses seleksi dilakukan secara terpisah untuk setiap wilayah sehingga kombinasi fitur yang dihasilkan bersifat adaptif terhadap kondisi lokal.

Tabel 4.3 Pola Fitur Terpilih pada Beberapa Wilayah

No	Wilayah	Jumlah Data	Jumlah Fitur	Fitur Terpilih
1	Bengkulu	1751	26	Kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin
2	Pekanbaru	1725	26	Kecepatan angin; titik embun; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi

				sebelumnya; rata-rata eksponensial kecepatan angin
3	Sleman	1584	26	Kelembapan; tekanan udara; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; rata-rata eksponensial kelembapan; komponen sinus hari dalam minggu
4	Malang	1706	26	Kelembapan; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kelembapan data observasi sebelumnya; rata-rata kelembapan dari 3 data observasi sebelumnya; rata-rata eksponensial kelembapan; rata-rata eksponensial titik embun
5	Medan	1821	26	Kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin; rata-rata eksponensial titik embun
6	Kemayoran	1779	26	Kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin

7	Banjarbaru	1819	26	Suhu udara; kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin
8	Samarinda	1766	26	Titik embun; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; titik embun data observasi sebelumnya; rata-rata titik embun dari 3 data observasi sebelumnya; rata-rata eksponensial titik embun
9	Jambi	1799	26	Kecepatan angin; titik embun; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin
10	Semarang	1583	26	Tekanan udara; kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin
11	Sorong	1216	26	PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; rata-rata kelembapan dari 3

				data observasi sebelumnya; rata-rata eksponensial kelembapan; tekanan udara data observasi sebelumnya; rata-rata tekanan udara dari 3 data observasi sebelumnya; rata-rata eksponensial tekanan udara
12	Kototabang	1395	26	PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; rata-rata eksponensial kelembapan; tekanan udara data observasi sebelumnya; rata-rata tekanan udara dari 3 data observasi sebelumnya; rata-rata eksponensial tekanan udara
13	Talang Betutu	1817	26	Kecepatan angin; PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; kecepatan angin data observasi sebelumnya; rata-rata kecepatan angin dari 3 data observasi sebelumnya; rata-rata eksponensial kecepatan angin
14	Pangkalanbun	1764	26	PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; perubahan PM2.5 dari data observasi sebelumnya; rata-rata eksponensial PM2.5; rata-rata eksponensial suhu udara; tekanan udara data observasi sebelumnya; rata-rata tekanan udara dari 3 data observasi sebelumnya; rata-rata eksponensial tekanan udara
15	Maros	1473	26	PM2.5 data observasi sebelumnya; rata-rata PM2.5 dari 3 data observasi sebelumnya; rata-rata eksponensial PM2.5; kelembapan data observasi

				sebelumnya; rata-rata kelembapan dari 3 data observasi sebelumnya; rata-rata eksponensial kelembapan; rata-rata titik embun dari 3 data observasi sebelumnya; rata-rata eksponensial titik embun
--	--	--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Berdasarkan hasil yang ditampilkan pada Tabel 4.3, terlihat bahwa fitur-fitur yang berkaitan dengan histori konsentrasi PM2.5 merupakan fitur yang paling dominan dan secara konsisten terpilih di seluruh wilayah penelitian. Fitur tersebut meliputi nilai PM2.5 pada data observasi sebelumnya, rata-rata PM2.5 dari tiga data observasi sebelumnya, perubahan PM2.5 terhadap observasi sebelumnya, serta rata-rata eksponensial PM2.5 dari serangkaian observasi sebelumnya. Temuan ini menunjukkan bahwa konsentrasi PM2.5 memiliki sifat persistensi yang kuat, di mana kondisi kualitas udara pada observasi sebelumnya berpengaruh signifikan terhadap konsentrasi PM2.5 pada observasi berikutnya. Dengan demikian, informasi historis PM2.5 menjadi komponen utama yang dapat dimanfaatkan secara efektif oleh model dalam proses peramalan.

Selain fitur berbasis PM2.5, sejumlah variabel meteorologi juga terpilih dan berkontribusi dalam pemodelan, seperti kecepatan angin, kelembapan udara, tekanan udara, suhu udara, serta titik embun. Namun, tingkat kepentingan variabel meteorologi tersebut tidak bersifat seragam di seluruh wilayah. Perbedaan fitur meteorologi yang terpilih mencerminkan variasi karakteristik lingkungan, kondisi atmosfer, serta faktor geografis lokal yang memengaruhi proses pembentukan, akumulasi, dan dispersi partikulat PM2.5 pada masing-masing wilayah penelitian.

Pada beberapa wilayah seperti Sorong, Kototabang, dan Maros, fitur meteorologi yang berkaitan dengan kelembapan, titik embun, atau tekanan udara relatif lebih dominan dibandingkan fitur angin. Sebaliknya, pada wilayah lain seperti Bengkulu, Pekanbaru, Medan, Kemayoran, Jambi, dan Semarang, fitur yang berkaitan dengan kecepatan angin lebih sering terpilih, yang mengindikasikan peran penting angin dalam proses dispersi polutan. Variasi ini menunjukkan bahwa tidak terdapat satu kombinasi fitur yang optimal untuk seluruh wilayah, sehingga pendekatan seleksi fitur yang adaptif menjadi sangat penting dalam membangun model peramalan yang *robust*.

Secara keseluruhan, penerapan seleksi fitur menggunakan *SelectKBest* tidak hanya berperan dalam mengurangi dimensi data dan meningkatkan efisiensi komputasi, tetapi juga membantu model *Support Vector Regression* untuk lebih fokus pada fitur-fitur yang paling relevan. Pendekatan ini berkontribusi dalam menjaga stabilitas kinerja model serta meningkatkan kemampuan generalisasi, meskipun dihadapkan pada perbedaan karakteristik data dan kondisi lingkungan antar wilayah penelitian.

#### **4.4 Implementasi Model *Support Vector Regression***

Pada tahap ini dilakukan implementasi model peramalan kualitas udara menggunakan metode *Support Vector Regression* (SVR) untuk meramalkan konsentrasi PM<sub>2.5</sub> satu hari ke depan (H+1). Implementasi model dilakukan secara terintegrasi dalam sebuah *pipeline* peramalan yang mencakup proses pembentukan dataset *time series*, pra-pemrosesan data, rekayasa fitur, pelatihan model, evaluasi kinerja, hingga penyajian hasil peramalan ke dalam sistem informasi berbasis web.

Pendekatan ini dirancang agar proses peramalan dapat berjalan secara otomatis dan konsisten untuk setiap wilayah penelitian.

Model SVR diimplementasikan menggunakan bahasa pemrograman Python dan dijalankan sebagai layanan terpisah berbasis Flask. Layanan ini menerima data historis yang telah diproses dari sistem *backend* Laravel, kemudian melakukan proses pelatihan dan pengujian model untuk masing-masing wilayah. Hasil peramalan yang dihasilkan berupa nilai peramalan PM2.5, metrik evaluasi model, serta informasi pendukung lainnya yang kemudian dikembalikan ke sistem Laravel untuk disimpan dan ditampilkan kepada pengguna.

Implementasi SVR dilakukan secara terpisah untuk setiap wilayah penelitian. Hal ini bertujuan untuk menangkap karakteristik lokal dari masing-masing wilayah, mengingat kondisi geografis, meteorologis, dan aktivitas antropogenik yang berbeda dapat memengaruhi pola kualitas udara. Dengan pendekatan ini, model peramalan tidak dipaksakan bersifat global, melainkan lebih adaptif terhadap kondisi spesifik setiap wilayah.

#### **4.4.1 Konfigurasi Model**

Model peramalan yang digunakan dalam penelitian ini adalah *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF). Pemilihan kernel RBF didasarkan pada kemampuannya dalam memodelkan hubungan *non-linear* antara variabel input dan target, yang umum ditemukan pada data kualitas udara. Selain itu, SVR dikenal memiliki performa yang relatif stabil pada dataset dengan ukuran kecil hingga menengah, sehingga sesuai dengan karakteristik data penelitian ini.

Parameter utama yang digunakan dalam model SVR meliputi parameter regularisasi ( $C$ ), parameter kernel ( $\gamma$ ), serta parameter epsilon ( $\epsilon$ ) pada fungsi *loss*. Parameter-parameter tersebut menentukan tingkat kompleksitas model, sensitivitas terhadap kesalahan peramalan, serta kemampuan model dalam menangani *noise* pada data. Dalam implementasinya, nilai parameter dipilih pada rentang moderat untuk menghindari *overfitting*, mengingat jumlah data historis yang terbatas pada beberapa wilayah.

Proses pelatihan model dilakukan menggunakan data historis, dengan target berupa nilai PM2.5 pada hari berikutnya. Setiap model dilatih menggunakan skema validasi *walk-forward validation* sehingga urutan temporal data tetap terjaga. Dengan konfigurasi ini, model SVR diharapkan mampu mempelajari pola temporal PM2.5 secara optimal tanpa melanggar asumsi kausalitas waktu.

#### 4.4.2 Skema Validasi

Proses pelatihan dan evaluasi model *Support Vector Regression* (SVR) pada penelitian ini dilakukan dengan memperhatikan karakteristik data yang bersifat runtun waktu (*time-series*). Oleh karena itu, metode validasi yang digunakan tidak dilakukan secara acak (*random split*), melainkan menggunakan pendekatan *walk-forward validation* yang mempertahankan urutan kronologis data. Pendekatan ini bertujuan untuk memastikan bahwa model hanya dilatih menggunakan data masa lalu dan diuji pada data masa depan, sehingga menyerupai kondisi peramalan yang sesungguhnya.

Dalam skema *walk-forward validation*, data historis dibagi ke dalam beberapa lipatan (*fold*) secara bertahap. Pada setiap iterasi, sebagian data awal



digunakan sebagai data latih, sementara sebagian data berikutnya digunakan sebagai data uji. Selain itu, diterapkan jeda waktu (*gap*) antara data latih dan data uji untuk mengurangi potensi kebocoran informasi (*data leakage*), terutama akibat penggunaan fitur *lag* dan fitur berbasis perataan waktu seperti *rolling mean* dan *exponential weighted moving average*. Dengan adanya jeda ini, informasi dari masa depan tidak secara langsung memengaruhi proses pelatihan model.

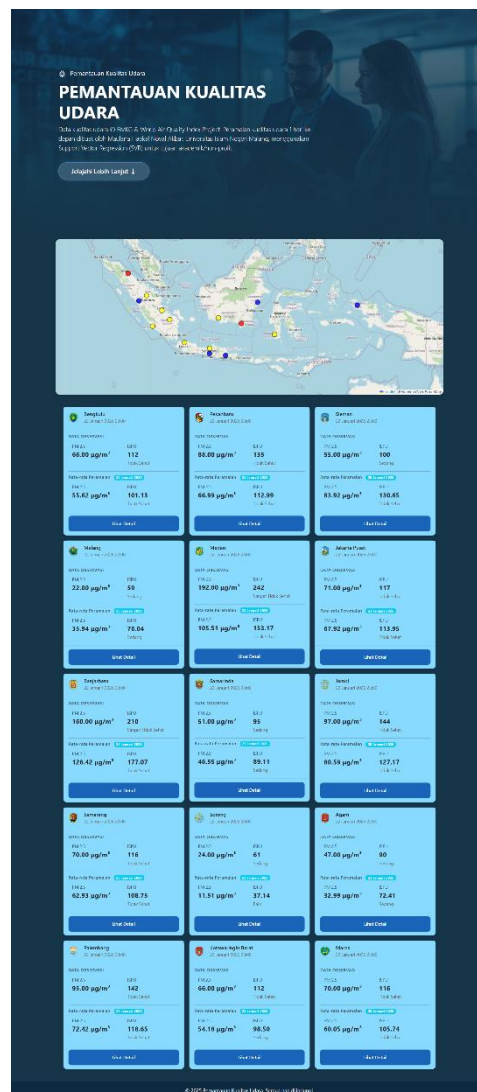
Proses pelatihan model dilakukan secara terpisah untuk setiap wilayah. Hal ini disebabkan oleh perbedaan karakteristik data antar wilayah, baik dari sisi jumlah data historis, tingkat fluktuasi PM2.5, maupun ketersediaan variabel meteorologi. Model SVR dilatih menggunakan fitur-fitur hasil normalisasi pada masing-masing wilayah, kemudian diuji untuk meramalkan nilai PM2.5 satu hari ke depan (H+1). Hasil peramalan tersebut selanjutnya digunakan sebagai dasar untuk menghitung indeks kualitas udara berdasarkan standar US EPA dan ISPU Indonesia.

Sebagai pembandingan kinerja model, digunakan model *baseline* berupa *persistence model*, yaitu model sederhana yang meramalkan nilai PM2.5 pada hari berikutnya sama dengan nilai PM2.5 pada hari sebelumnya. Model *baseline* ini digunakan untuk memberikan gambaran kinerja minimum yang diharapkan dari sebuah model peramalan runtun waktu. Dengan membandingkan hasil SVR terhadap *baseline*, dapat diketahui apakah model *machine learning* yang diterapkan benar-benar memberikan peningkatan kinerja dibandingkan pendekatan sederhana.

#### **4.5 Implementasi Sistem Informasi Berbasis Web**

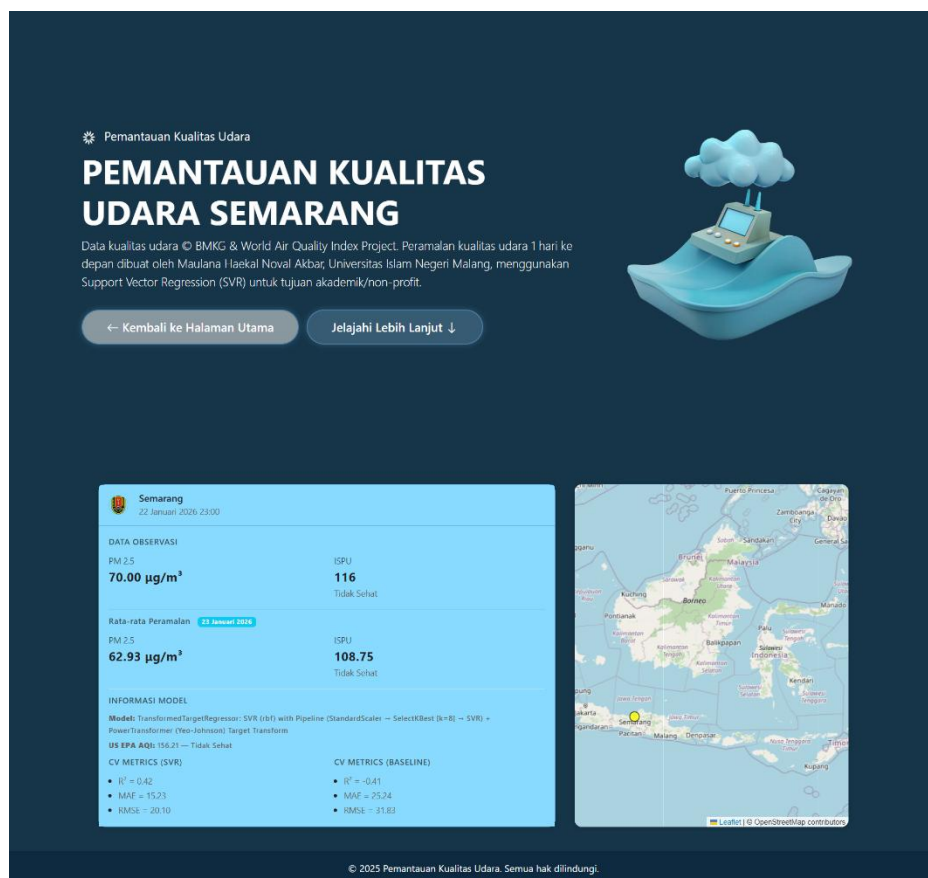
Sistem informasi pemantauan dan peramalan kualitas udara pada penelitian ini diimplementasikan dalam bentuk aplikasi berbasis web yang bertujuan untuk

menyajikan informasi kualitas udara secara visual, ringkas, dan mudah dipahami oleh pengguna. Sistem ini mengintegrasikan proses pengumpulan data kualitas udara dari WAQI API, pemrosesan peramalan menggunakan model *Support Vector Regression* (SVR), serta penyajian hasil peramalan dalam satu platform terpadu. Dengan pendekatan ini, pengguna tidak hanya memperoleh informasi kondisi udara terkini, tetapi juga gambaran kondisi kualitas udara untuk satu hari ke depan (H+1). Implementasi sistem berbasis web memungkinkan akses informasi secara luas tanpa memerlukan instalasi aplikasi tambahan.



Gambar 4.1 Halaman Utama Website

Gambar 4.1 menampilkan halaman utama sistem informasi kualitas udara. Pada halaman ini disajikan peta interaktif wilayah Indonesia yang dilengkapi dengan penanda berwarna sesuai kategori Indeks Standar Pencemar Udara (ISPU). Setiap penanda mewakili satu wilayah pemantauan dan menggunakan skema warna untuk menunjukkan tingkat kualitas udara, sehingga pengguna dapat dengan cepat mengidentifikasi wilayah dengan kondisi udara baik maupun tidak sehat. Selain peta, halaman utama juga menampilkan tabel ringkasan yang memuat informasi nama wilayah, tanggal pengamatan, nilai PM2.5 aktual, kategori ISPU, serta hasil peramalan PM2.5 dan ISPU untuk satu hari ke depan. Apabila suatu wilayah belum memiliki hasil peramalan, sistem secara otomatis menampilkan keterangan bahwa data peramalan belum tersedia.



Gambar 4.2 Halaman Detail Wilayah Website

Gambar 4.2 menunjukkan halaman detail untuk salah satu wilayah pemantauan. Halaman ini dirancang untuk memberikan informasi yang lebih lengkap dan spesifik terkait kualitas udara di wilayah tersebut. Informasi yang ditampilkan meliputi nilai PM2.5 aktual, hasil peramalan PM2.5 (H+1), kategori kualitas udara berdasarkan standar US EPA dan ISPU Indonesia, serta metrik evaluasi model berupa MAE, RMSE, dan  $R^2$ . Penyajian metrik evaluasi ini bertujuan untuk memberikan transparansi kepada pengguna mengenai tingkat keandalan hasil peramalan yang dihasilkan oleh model SVR. Dengan demikian, pengguna dapat memahami bahwa hasil peramalan memiliki tingkat ketidakpastian tertentu.

Secara keseluruhan, implementasi sistem informasi berbasis web dalam penelitian ini telah berhasil mengintegrasikan aspek pemantauan kualitas udara, peramalan jangka pendek, dan penyajian informasi kepada pengguna. Sistem ini tidak hanya berfungsi sebagai media visualisasi hasil peramalan, tetapi juga sebagai sarana pendukung pengambilan keputusan berbasis data kualitas udara. Keberadaan sistem ini memperkuat kontribusi penelitian, tidak hanya dari sisi pengembangan model peramalan, tetapi juga dari sisi penerapan nyata dalam bentuk sistem informasi.

#### **4.6 Hasil Evaluasi Model**

Evaluasi model peramalan kualitas udara dilakukan untuk menilai sejauh mana model *Support Vector Regression* (SVR) mampu meramalkan konsentrasi PM2.5 satu hari ke depan (H+1) berdasarkan data historis yang tersedia. Proses

evaluasi ini menjadi tahap penting karena hasil peramalan akan digunakan sebagai dasar penentuan indeks kualitas udara, baik berdasarkan standar US EPA maupun ISPU Indonesia. Dengan demikian, keandalan model perlu diukur secara kuantitatif agar dapat diketahui tingkat akurasi serta keterbatasannya.

Evaluasi dilakukan secara terpisah untuk setiap wilayah penelitian, mengingat perbedaan karakteristik data antar wilayah, seperti jumlah data historis dan tingkat fluktuasi PM<sub>2.5</sub>. Hasil evaluasi tidak hanya digunakan untuk menilai kinerja absolut model SVR, tetapi juga untuk membandingkannya dengan model *baseline* berupa *persistence model*. Perbandingan ini bertujuan untuk memastikan bahwa penggunaan metode *machine learning* memberikan peningkatan kinerja dibandingkan pendekatan sederhana yang umum digunakan pada peramalan runtun waktu.

#### 4.6.1 Metrik Evaluasi

Evaluasi kinerja model peramalan kualitas udara dalam penelitian ini dilakukan menggunakan tiga metrik statistik, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi ( $R^2$ ). Pemilihan ketiga metrik tersebut bertujuan untuk memberikan gambaran yang komprehensif mengenai tingkat kesalahan peramalan serta kemampuan model dalam menjelaskan variasi data PM<sub>2.5</sub>. Setiap metrik memiliki karakteristik yang berbeda sehingga saling melengkapi dalam menilai performa model. Evaluasi dilakukan menggunakan skema *walk-forward validation* agar tetap menjaga urutan temporal data dan menghindari kebocoran informasi dari data masa depan. Dengan

pendekatan ini, hasil evaluasi diharapkan dapat merepresentasikan kondisi kinerja model yang lebih realistis.

*Mean Absolute Error* (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai peramalan dan nilai aktual PM2.5. Metrik ini memberikan informasi yang mudah diinterpretasikan karena memiliki satuan yang sama dengan data asli, yaitu mikrogram per meter kubik ( $\mu\text{g}/\text{m}^3$ ). Nilai MAE yang lebih kecil menunjukkan bahwa hasil peramalan model semakin mendekati nilai aktual. MAE juga relatif tidak terlalu sensitif terhadap nilai ekstrem, sehingga cocok digunakan pada data kualitas udara yang cenderung memiliki fluktuasi harian. Dalam penelitian ini, MAE digunakan sebagai indikator utama tingkat kesalahan peramalan rata-rata model.

*Root Mean Squared Error* (RMSE) digunakan untuk mengukur akar dari rata-rata kuadrat kesalahan peramalan. Berbeda dengan MAE, RMSE memberikan penalti yang lebih besar terhadap kesalahan peramalan yang bernilai besar. Hal ini menjadikan RMSE sangat sensitif terhadap adanya lonjakan atau fluktuasi ekstrem pada konsentrasi PM2.5. Penggunaan RMSE penting dalam konteks kualitas udara karena kesalahan besar pada hari dengan polusi tinggi dapat berdampak signifikan terhadap interpretasi kondisi udara. Oleh karena itu, RMSE digunakan untuk mengevaluasi stabilitas dan konsistensi model dalam menangani variasi data yang ekstrem.

Koefisien determinasi ( $R^2$ ) digunakan untuk mengukur sejauh mana variasi nilai PM2.5 aktual dapat dijelaskan oleh model peramalan. Nilai  $R^2$  berada pada rentang negatif hingga satu, di mana nilai mendekati satu menunjukkan kemampuan

peramalan yang sangat baik. Nilai  $R^2$  yang mendekati nol menunjukkan bahwa model memiliki kemampuan yang setara dengan model rata-rata, sedangkan nilai negatif mengindikasikan bahwa model berkinerja lebih buruk dibandingkan model *baseline* sederhana. Dalam penelitian ini, nilai  $R^2$  digunakan sebagai indikator kemampuan model dalam menangkap pola temporal PM2.5. Namun demikian, interpretasi nilai  $R^2$  dilakukan secara hati-hati mengingat keterbatasan jumlah data dan resolusi temporal yang digunakan.

#### 4.6.2 Hasil Evaluasi Model SVR Per Wilayah dan Perbandingan dengan *Baseline*

Hasil evaluasi kinerja model *Support Vector Regression* (SVR) pada setiap wilayah penelitian dalam meramalkan konsentrasi PM2.5 satu hari ke depan (H+1) dianalisis menggunakan koefisien determinasi ( $R^2$ ), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE). Evaluasi ini bertujuan untuk mengamati variasi performa model antar wilayah sekaligus membandingkannya dengan model *baseline* sederhana yang menggunakan nilai PM2.5 periode sebelumnya sebagai peramalan. Rekapitulasi nilai rata-rata  $R^2$ , MAE, dan RMSE dari model SVR dan *baseline* pada seluruh wilayah penelitian ditampilkan pada Tabel 4.4.

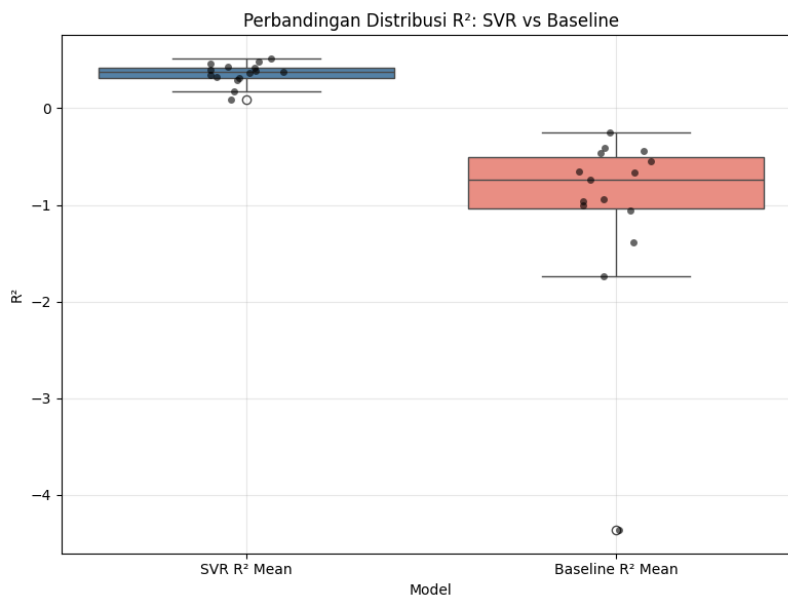
Tabel 4.4 Hasil Evaluasi Model SVR Per Wilayah dan Perbandingan dengan *Baseline*

No.	Wilayah	SVR $R^2$ <i>Mean</i>	SVR MAE <i>Mean</i>	SVR RMSE <i>Mean</i>	<i>Baseline</i> $R^2$ <i>Mean</i>	<i>Baseline</i> MAE <i>Mean</i>	<i>Baseline</i> RMSE <i>Mean</i>
1	Bengkulu	0.430	19.55	27.58	-0.253	32.21	41.30
2	Pekanbaru	0.377	18.99	26.35	-0.655	34.73	41.82
3	Sleman	0.370	17.14	23.12	-1.733	37.92	43.99
4	Malang	0.403	16.76	23.64	-1.004	35.89	43.42
5	Medan	0.512	21.20	28.76	-1.062	45.36	56.95
6	Kemayoran	0.463	16.53	21.75	-0.459	29.97	36.37
7	Banjarbaru	0.392	12.98	17.98	-0.668	22.15	28.95
8	Samarinda	0.288	12.72	16.80	-0.941	22.63	27.07

9	Jambi	0.321	18.29	25.85	-1.387	36.49	43.53
10	Semarang	0.418	15.23	20.10	-0.407	25.24	31.83
11	Sorong	0.089	8.69	12.97	-0.546	12.85	16.39
12	Kototabang	0.488	8.91	12.33	-0.743	17.24	22.12
13	Talang Betutu	0.312	16.50	22.96	-4.361	51.70	58.05
14	Pangkalanbun	0.341	10.57	14.42	-0.963	21.17	25.23
15	Maros	0.172	13.18	19.53	-0.437	17.97	24.80

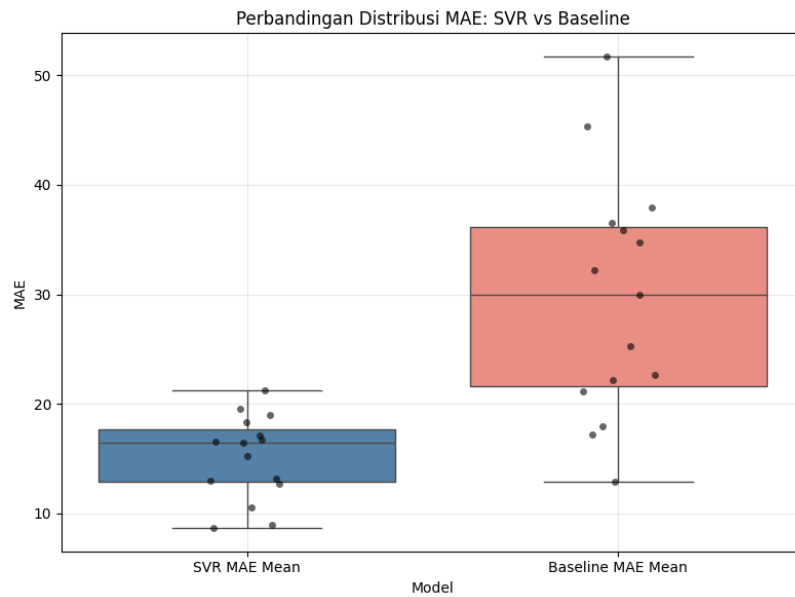
Berdasarkan Tabel 4.4, nilai MAE dan RMSE model *Support Vector Regression* (SVR) menunjukkan adanya variasi performa antar wilayah penelitian. Wilayah Kototabang dan Sorong memiliki nilai kesalahan yang relatif paling rendah dibandingkan wilayah lainnya. Kototabang memiliki MAE sebesar 8.91 dan RMSE sebesar 12.33, sedangkan Sorong memiliki MAE sebesar 8.69 dan RMSE sebesar 12.97. Nilai kesalahan yang rendah ini menunjukkan bahwa pola historis PM<sub>2.5</sub> serta variabel meteorologi di wilayah tersebut cenderung lebih stabil sehingga hubungan antar variabel dapat dipelajari dengan lebih baik oleh model. Sebaliknya, beberapa wilayah seperti Medan dan Bengkulu menunjukkan nilai kesalahan yang relatif lebih tinggi. Medan memiliki MAE sebesar 21.20 dan RMSE sebesar 28.76, sedangkan Bengkulu memiliki MAE sebesar 19.55 dan RMSE sebesar 27.58. Kondisi ini mengindikasikan adanya fluktuasi konsentrasi PM<sub>2.5</sub> dan variabilitas meteorologi yang lebih kompleks sehingga lebih sulit dimodelkan. Kemampuan model dalam menjelaskan variasi data dianalisis melalui koefisien determinasi ( $R^2$ ), sebagaimana divisualisasikan pada Gambar 4.3.





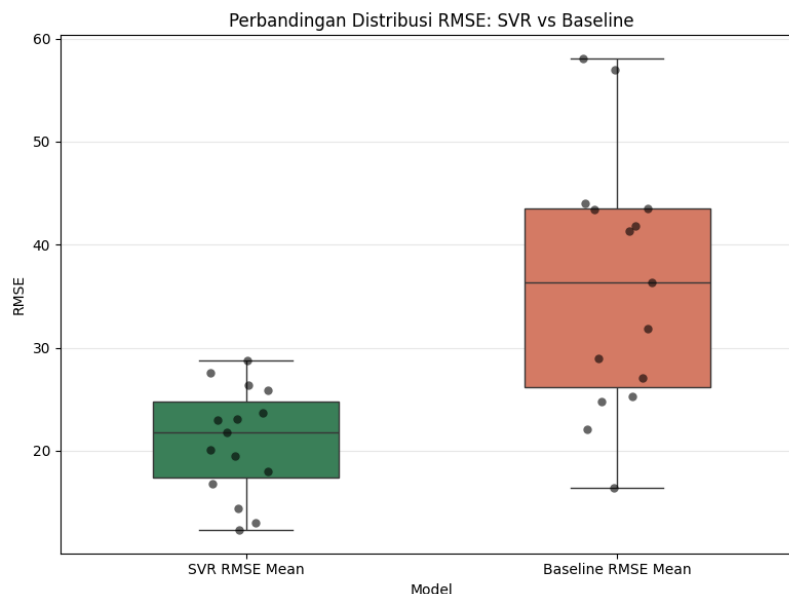
Gambar 4.3 Perbandingan Distribusi R<sup>2</sup>: SVR vs *Baseline*

Berdasarkan Gambar 4.3, distribusi nilai R<sup>2</sup> model SVR terlihat lebih tinggi dibandingkan model *baseline*. Nilai R<sup>2</sup> model SVR berada pada rentang 0.089 hingga 0.512, dengan nilai tertinggi terdapat pada wilayah Medan (0.512), diikuti Kototabang (0.488) dan Kemayoran (0.463). Hal ini menunjukkan bahwa model SVR mampu menjelaskan sebagian variasi data PM<sub>2.5</sub> pada berbagai wilayah penelitian. Sebaliknya, model *baseline* memiliki nilai R<sup>2</sup> yang seluruhnya negatif, bahkan pada wilayah Talang Betutu mencapai -4.361. Nilai R<sup>2</sup> negatif menunjukkan bahwa model *baseline* tidak mampu menjelaskan variasi data secara baik dan memiliki performa yang lebih buruk dibandingkan model SVR. Perbandingan *Mean Absolute Error* (MAE) antara model SVR dan *baseline* ditunjukkan pada Gambar 4.4.



Gambar 4.4 Perbandingan Distribusi MAE: SVR vs *Baseline*

Berdasarkan Gambar 4.4, distribusi nilai MAE model SVR lebih rendah dibandingkan model *baseline*. Nilai MAE model SVR berada pada rentang 8.69 hingga 21.20, sedangkan *baseline* memiliki nilai MAE yang lebih tinggi yaitu antara 12.85 hingga 51.70. Perbedaan ini menunjukkan bahwa model SVR menghasilkan kesalahan peramalan yang lebih kecil dibandingkan model *baseline* sehingga lebih akurat dalam meramalkan konsentrasi PM<sub>2.5</sub>. Perbandingan *Root Mean Squared Error* (RMSE) antara model SVR dan *baseline* akan ditunjukkan pada Gambar 4.5.



Gambar 4.5 Perbandingan Distribusi RMSE: SVR vs *Baseline*

Distribusi nilai RMSE pada Gambar 4.5 juga menunjukkan pola yang serupa. Nilai RMSE model SVR berada pada rentang 12.33 hingga 28.76, sedangkan *baseline* memiliki nilai yang lebih tinggi yaitu antara 16.39 hingga 58.05. Hal ini menunjukkan bahwa model SVR lebih efektif dalam meminimalkan kesalahan peramalan dibandingkan model *baseline*.

Nilai  $R^2$  pada beberapa wilayah berada pada kategori rendah hingga sedang. Kondisi ini masih wajar dalam peramalan kualitas udara karena konsentrasi PM<sub>2.5</sub> dipengaruhi oleh banyak faktor lingkungan yang kompleks dan dinamis. Konsentrasi PM<sub>2.5</sub> dipengaruhi oleh berbagai faktor seperti kondisi meteorologi, aktivitas manusia, serta proses atmosfer yang tidak seluruhnya dapat direpresentasikan dalam model. Selain itu, peramalan yang dilakukan bersifat jangka pendek yaitu satu hari ke depan (H+1), sehingga perubahan kondisi lingkungan yang terjadi secara cepat tetap sulit diramalkan secara presisi. Oleh

karena itu, nilai  $R^2$  pada rentang tersebut masih dianggap wajar dalam studi peramalan kualitas udara.

Namun demikian, hasil evaluasi menunjukkan bahwa model SVR secara konsisten memiliki performa yang lebih baik dibandingkan model *baseline* pada seluruh wilayah penelitian. Hal ini terlihat dari nilai MAE dan RMSE model SVR yang lebih rendah dibandingkan *baseline* pada setiap lokasi, serta nilai  $R^2$  yang seluruhnya berada pada rentang positif. Sebaliknya, model *baseline* menunjukkan nilai  $R^2$  negatif pada semua wilayah, yang mengindikasikan bahwa model tersebut tidak mampu menjelaskan variasi data secara memadai. Dengan demikian, berdasarkan kombinasi metrik evaluasi  $R^2$ , MAE, dan RMSE, dapat disimpulkan bahwa model SVR lebih efektif dalam menangkap pola variasi data dan menghasilkan peramalan konsentrasi PM<sub>2.5</sub> yang lebih akurat dibandingkan metode *baseline*.

#### 4.6.3 Hasil Eksperimen Terbatas

Subbab ini menyajikan hasil eksperimen terbatas yang dilakukan untuk mengevaluasi sensitivitas dan batasan model *Support Vector Regression* (SVR) dalam meramalkan konsentrasi PM<sub>2.5</sub> satu hari ke depan (H+1). Eksperimen terbatas dilakukan sebagai bagian dari analisis tambahan untuk mengetahui pengaruh variasi konfigurasi fitur, strategi pembobotan waktu, seleksi fitur, serta transformasi target terhadap performa model.

Berbeda dengan evaluasi utama pada Subbab 4.6.2 yang menggunakan konfigurasi standar model, eksperimen terbatas ini dirancang untuk menguji beberapa skenario alternatif yang umum digunakan dalam peramalan runtun waktu.

Setiap eksperimen dilakukan pada seluruh wilayah penelitian sehingga dapat diketahui apakah suatu pendekatan memberikan peningkatan kinerja secara konsisten atau hanya efektif pada wilayah tertentu.

Hasil eksperimen terbatas ini tidak hanya berfungsi sebagai pembandingan teknis, tetapi juga memberikan gambaran mengenai faktor-faktor yang memengaruhi keberhasilan model SVR dalam menangkap pola temporal PM2.5. Dengan demikian, subbab ini menjadi dasar dalam menjelaskan keterbatasan penelitian sekaligus peluang pengembangan pada penelitian selanjutnya.

#### **4.6.3.1. Eksperimen 1: Variasi Panjang *Lag* dan *Rolling Window***

Eksperimen pertama dilakukan untuk menguji pengaruh variasi panjang *lag* dan *rolling window* dalam proses rekayasa fitur terhadap performa model *Support Vector Regression* (SVR). Pada peramalan runtun waktu, fitur *lag* digunakan untuk merepresentasikan nilai historis PM2.5 pada data observasi sebelumnya, sedangkan *rolling window* digunakan untuk menangkap kecenderungan rata-rata bergerak dalam periode tertentu. Pemilihan kombinasi *lag* dan *rolling window* yang tepat diharapkan dapat meningkatkan kemampuan model dalam mempelajari pola temporal serta mengurangi pengaruh fluktuasi ekstrem pada data PM2.5 harian.

Pada eksperimen ini, tiga konfigurasi diuji pada setiap wilayah penelitian, yaitu *lag* 1 dengan *rolling window* 3, *lag* 3 dengan *rolling window* 7, serta *lag* 7 dengan *rolling window* 14. Kinerja masing-masing konfigurasi dievaluasi menggunakan metrik koefisien determinasi ( $R^2$ ), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE). Berdasarkan hasil evaluasi tersebut, konfigurasi terbaik untuk setiap wilayah ditentukan menggunakan nilai  $R^2$  tertinggi

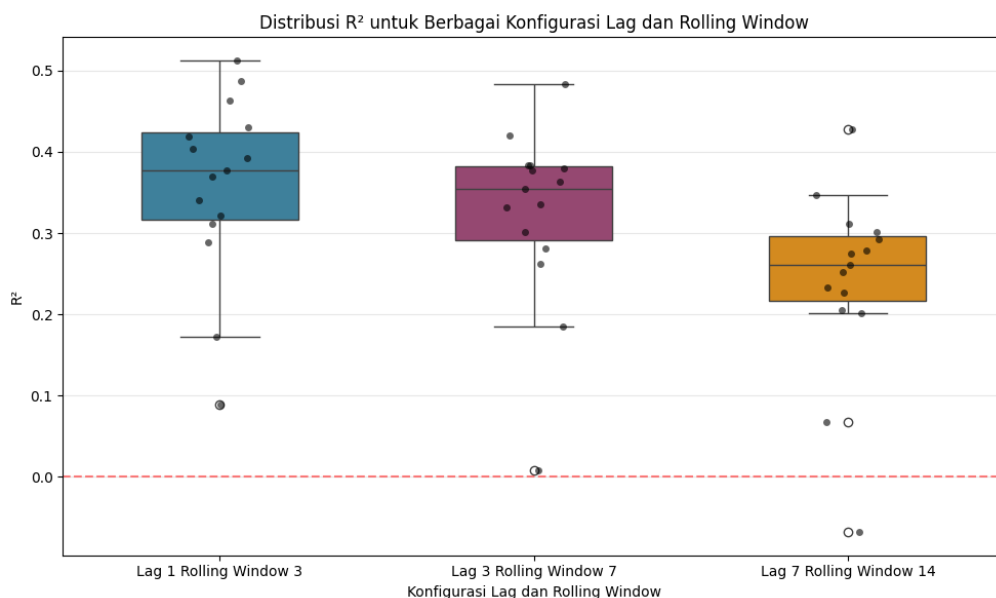
sebagai indikator utama kemampuan model dalam menjelaskan variasi data PM2.5. Ringkasan hasil eksperimen yang menampilkan konfigurasi terbaik pada setiap wilayah beserta nilai rata-rata  $R^2$ , MAE, dan RMSE disajikan pada Tabel 4.5. Tabel ini hanya menampilkan konfigurasi terbaik pada setiap wilayah, sedangkan perbandingan lengkap antar konfigurasi disajikan dalam bentuk visualisasi pada Gambar 4.6–4.8.

Tabel 4.5 Ringkasan Konfigurasi Terbaik pada Eksperimen 1: Variasi *Lag* dan *Rolling Window*

<b>Wilayah</b>	<b>Konfigurasi Terbaik</b>	<b><math>R^2</math> <i>Mean</i></b>	<b>MAE <i>Mean</i></b>	<b>RMSE <i>Mean</i></b>
Bengkulu	<i>Lag 1 Rolling Window 3</i>	0.4300	19.5540	27.5801
Pekanbaru	<i>Lag 3 Rolling Window 7</i>	0.3840	19.4023	26.3073
Sleman	<i>Lag 1 Rolling Window 3</i>	0.3697	17.1405	23.1198
Malang	<i>Lag 1 Rolling Window 3</i>	0.4033	16.7581	23.6372
Medan	<i>Lag 1 Rolling Window 3</i>	0.5120	21.2028	28.7640
Kemayoran	<i>Lag 1 Rolling Window 3</i>	0.4626	16.5324	21.7465
Banjarbaru	<i>Lag 1 Rolling Window 3</i>	0.3923	12.9842	17.9793
Samarinda	<i>Lag 1 Rolling Window 3</i>	0.2883	12.7232	16.7975
Jambi	<i>Lag 1 Rolling Window 3</i>	0.3213	18.2876	25.8464
Semarang	<i>Lag 1 Rolling Window 3</i>	0.4183	15.2253	20.0996
Sorong	<i>Lag 1 Rolling Window 3</i>	0.0887	8.6948	12.9730
Kototabang	<i>Lag 1 Rolling Window 3</i>	0.4876	8.9113	12.3306
Talang Betutu	<i>Lag 1 Rolling Window 3</i>	0.3119	16.4952	22.9591
Pangkalanbun	<i>Lag 1 Rolling Window 3</i>	0.3410	10.5683	14.4178
Maros	<i>Lag 3 Rolling Window 7</i>	0.1847	13.4043	19.4335

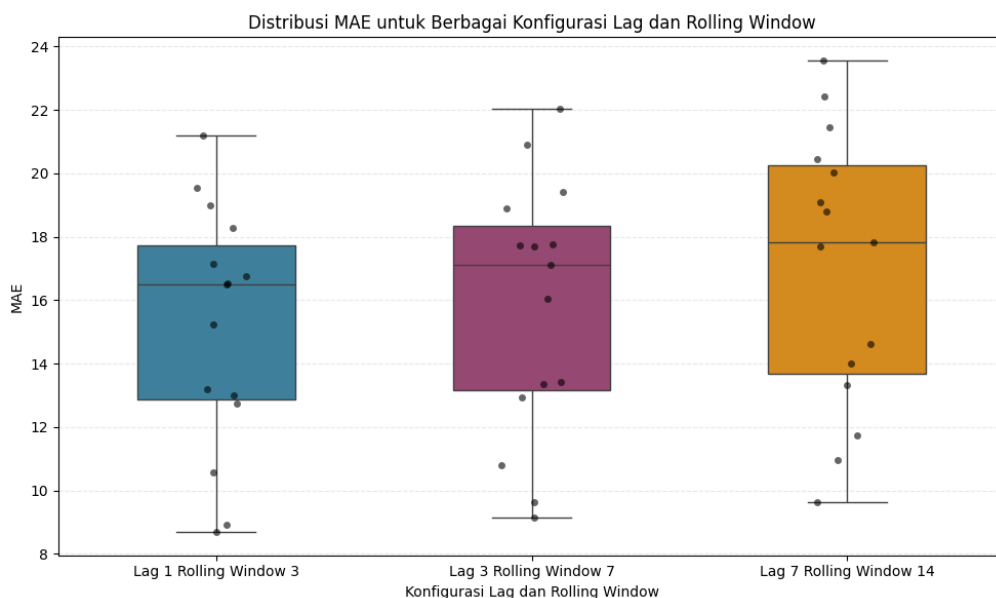
Berdasarkan Tabel 4.5, hampir seluruh wilayah menunjukkan konfigurasi terbaik pada *lag 1* dan *rolling window 3*, sedangkan Pekanbaru dan Maros lebih optimal pada *lag 3* dan *rolling window 7*. Wilayah seperti Medan ( $R^2 = 0.5120$ ), Kototabang ( $R^2 = 0.4876$ ), Kemayoran ( $R^2 = 0.4626$ ), dan Bengkulu ( $R^2 = 0.4300$ ) memiliki nilai  $R^2$  relatif lebih tinggi dibandingkan wilayah lainnya. Hal ini menunjukkan bahwa model SVR lebih mampu menjelaskan variasi PM2.5 ketika memanfaatkan informasi historis jangka pendek. Temuan ini menunjukkan bahwa

pola konsentrasi PM2.5 memiliki ketergantungan jangka pendek (*short-term dependency*), sehingga informasi terbaru lebih relevan dibandingkan data historis yang terlalu panjang dalam proses peramalan. Di sisi lain, Sorong memiliki nilai  $R^2$  paling rendah pada konfigurasi terbaiknya yaitu 0.0887, meskipun nilai MAE sebesar 8.6948 dan RMSE sebesar 12.9730 termasuk rendah. Kondisi ini mengindikasikan bahwa variasi PM2.5 di wilayah tersebut memiliki variasi yang relatif kecil atau pola yang lebih stabil, sehingga meskipun kesalahan peramalan rendah, kemampuan model dalam menjelaskan keragaman data tetap terbatas. Untuk memperjelas perbandingan performa antar konfigurasi *lag* dan *rolling window*, nilai  $R^2$  dari setiap konfigurasi divisualisasikan pada Gambar 4.6. Visualisasi ini digunakan untuk menunjukkan pengaruh variasi panjang *lag* dan *rolling window* terhadap kemampuan model dalam menjelaskan variasi data PM2.5 pada setiap wilayah penelitian.



Gambar 4.6 Perbandingan Distribusi Nilai  $R^2$  Model SVR pada Konfigurasi Panjang *Lag* dan *Rolling Window*

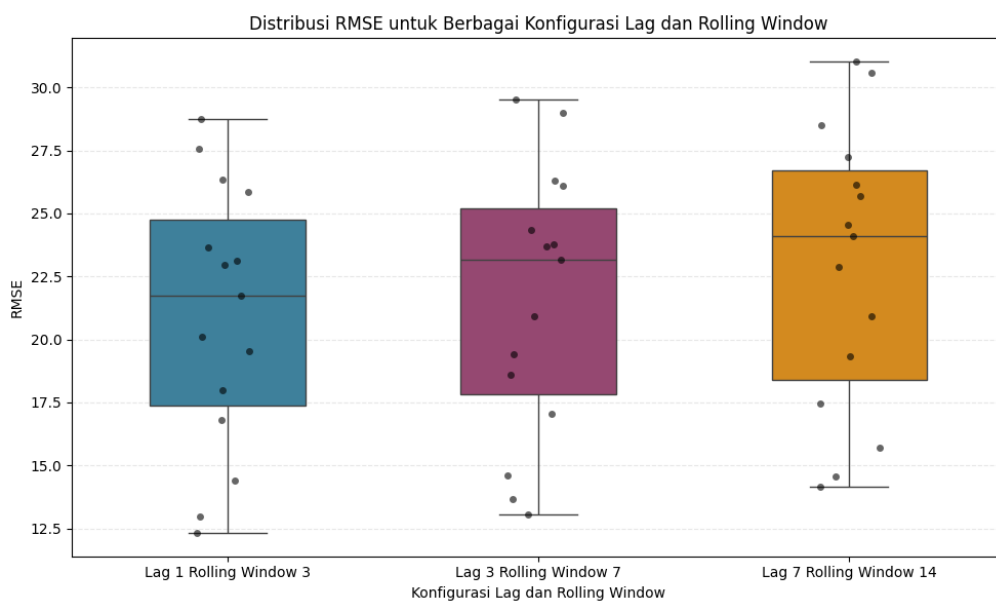
Berdasarkan Gambar 4.6, distribusi nilai  $R^2$  menunjukkan bahwa konfigurasi *Lag 1 Rolling Window 3* secara umum menghasilkan performa terbaik pada sebagian besar wilayah. Pada beberapa wilayah, seperti Bengkulu, Sleman, Malang, Medan, Kemayoran, Banjarbaru, Samarinda, Jambi, Semarang, Sorong, Kototabang, Talang Betutu, dan Pangkalanbun, konfigurasi ini memberikan nilai  $R^2$  tertinggi dibandingkan konfigurasi lainnya. Namun, terdapat pengecualian pada Pekanbaru dan Maros, di mana konfigurasi *Lag 3 Rolling Window 7* memberikan nilai  $R^2$  yang sedikit lebih tinggi. Sebaliknya, konfigurasi *Lag 7 Rolling Window 14* cenderung menghasilkan nilai  $R^2$  terendah di seluruh wilayah, yang menunjukkan bahwa penggunaan *lag* yang terlalu panjang justru menurunkan kemampuan model dalam menangkap pola data. Dengan demikian, pemanfaatan informasi historis jangka pendek lebih efektif dibandingkan penggunaan informasi historis yang terlalu panjang pada penelitian ini. Metrik kesalahan peramalan seperti MAE untuk setiap konfigurasi ditunjukkan lebih lanjut pada Gambar 4.7.



Gambar 4.7 Perbandingan Distribusi Nilai MAE Model SVR pada Konfigurasi Panjang *Lag* dan *Rolling Window*



Berdasarkan Gambar 4.7, distribusi nilai MAE menunjukkan pola yang konsisten dengan hasil  $R^2$ . Konfigurasi *Lag 1 Rolling Window 3* umumnya menghasilkan nilai MAE yang lebih rendah dibandingkan konfigurasi lainnya, yang menandakan bahwa kesalahan peramalan model lebih kecil pada konfigurasi tersebut. Pada beberapa wilayah, konfigurasi *Lag 3 Rolling Window 7* masih memberikan hasil yang kompetitif, bahkan menjadi konfigurasi terbaik di Pekanbaru dan Maros. Sementara itu, konfigurasi *Lag 7 Rolling Window 14* cenderung menghasilkan nilai MAE yang lebih tinggi, menunjukkan bahwa penambahan panjang *lag* dan *rolling window* tidak selalu meningkatkan performa model. Hal ini menunjukkan bahwa semakin panjang rentang historis yang digunakan, semakin besar kemungkinan model menangkap *noise* atau pola yang kurang relevan terhadap peramalan jangka pendek. Sebagai pelengkap, distribusi nilai RMSE untuk setiap konfigurasi ditunjukkan pada Gambar 4.8.



Gambar 4.8 Perbandingan Distribusi Nilai RMSE Model SVR pada Konfigurasi Panjang *Lag* dan *Rolling Window*

Berdasarkan Gambar 4.8, distribusi nilai RMSE menunjukkan pola yang serupa dengan MAE. Konfigurasi *Lag 1 Rolling Window 3* cenderung menghasilkan nilai RMSE yang lebih rendah dan lebih stabil dibandingkan konfigurasi lainnya, sehingga menunjukkan kemampuan model yang lebih baik dalam meminimalkan kesalahan peramalan. Pada wilayah Pekanbaru dan Maros, konfigurasi *Lag 3 Rolling Window 7* memberikan hasil yang sedikit lebih baik dibandingkan *Lag 1 Rolling Window 3*, tetapi secara umum perbedaannya tidak terlalu besar. Sebaliknya, konfigurasi *Lag 7 Rolling Window 14* hampir selalu menghasilkan nilai RMSE tertinggi, yang menandakan bahwa penggunaan *lag* yang terlalu panjang justru menurunkan akurasi model.

Secara keseluruhan, hasil Eksperimen 1 menunjukkan bahwa variasi panjang *lag* dan *rolling window* mempengaruhi performa model SVR dalam meramalkan konsentrasi PM<sub>2.5</sub>. Konfigurasi *lag 1* dan *rolling window 3* menjadi konfigurasi terbaik pada sebagian besar wilayah penelitian. Konfigurasi tersebut menghasilkan nilai R<sup>2</sup> yang lebih tinggi serta nilai MAE dan RMSE yang lebih rendah dibandingkan konfigurasi lainnya. Meskipun pada beberapa wilayah seperti Pekanbaru dan Maros konfigurasi *lag 3* dan *rolling window 7* memberikan hasil yang sedikit lebih baik, perbedaannya tidak terlalu signifikan. Oleh karena itu, konfigurasi *lag 1* dan *rolling window 3* dipilih sebagai dasar rekayasa fitur pada eksperimen selanjutnya karena memberikan performa yang paling stabil. Pemilihan konfigurasi ini juga menunjukkan bahwa penggunaan fitur yang sederhana namun relevan terhadap karakteristik data mampu memberikan performa yang optimal tanpa meningkatkan kompleksitas model secara signifikan.

#### 4.6.3.2. Eksperimen 2: Pengaruh Pembobotan Waktu Eksponensial

Eksperimen kedua bertujuan untuk menganalisis pengaruh penerapan pembobotan waktu eksponensial (*exponential time-decay weighting*) terhadap performa model *Support Vector Regression* (SVR) dalam peramalan konsentrasi PM<sub>2.5</sub> harian. Dalam peramalan runtun waktu, pembobotan waktu digunakan untuk memberikan bobot yang lebih besar pada data terbaru dibandingkan data historis yang lebih lama. Pendekatan ini didasarkan pada asumsi bahwa kondisi kualitas udara terkini memiliki pengaruh yang lebih kuat dalam meramalkan konsentrasi PM<sub>2.5</sub> pada hari berikutnya.

Eksperimen ini membandingkan dua skenario utama, yaitu model SVR tanpa pembobotan dan model SVR dengan pembobotan waktu eksponensial. Pada skenario pembobotan, digunakan lima variasi nilai *half-life*, yaitu 60, 90, 120, 150, dan 180 hari. Variasi ini bertujuan untuk mengevaluasi sensitivitas performa model terhadap tingkat peluruhan bobot terhadap waktu. Nilai *half-life* yang lebih kecil (misalnya 60 hari) menyebabkan peluruhan bobot berlangsung lebih cepat sehingga model lebih menekankan data terbaru. Sebaliknya, nilai *half-life* yang lebih besar (misalnya 180 hari) menghasilkan peluruhan bobot yang lebih lambat sehingga data historis tetap memiliki kontribusi yang signifikan dalam proses pelatihan.

Seluruh skenario diuji menggunakan prosedur hyperparameter tuning yang sama untuk menjaga konsistensi dan keadilan perbandingan. Evaluasi performa dilakukan pada seluruh wilayah penelitian menggunakan tiga metrik, yaitu koefisien determinasi ( $R^2$ ), *Mean Absolute Error* (MAE), dan *Root Mean Squared*

*Error* (RMSE). Ringkasan hasil eksperimen yang menampilkan konfigurasi terbaik pada setiap wilayah beserta nilai rata-rata  $R^2$ , MAE, dan RMSE disajikan pada Tabel 4.6. Tabel ini hanya menampilkan konfigurasi terbaik pada setiap wilayah, sedangkan perbandingan lengkap antar konfigurasi disajikan dalam bentuk visualisasi pada Gambar 4.9–4.11.

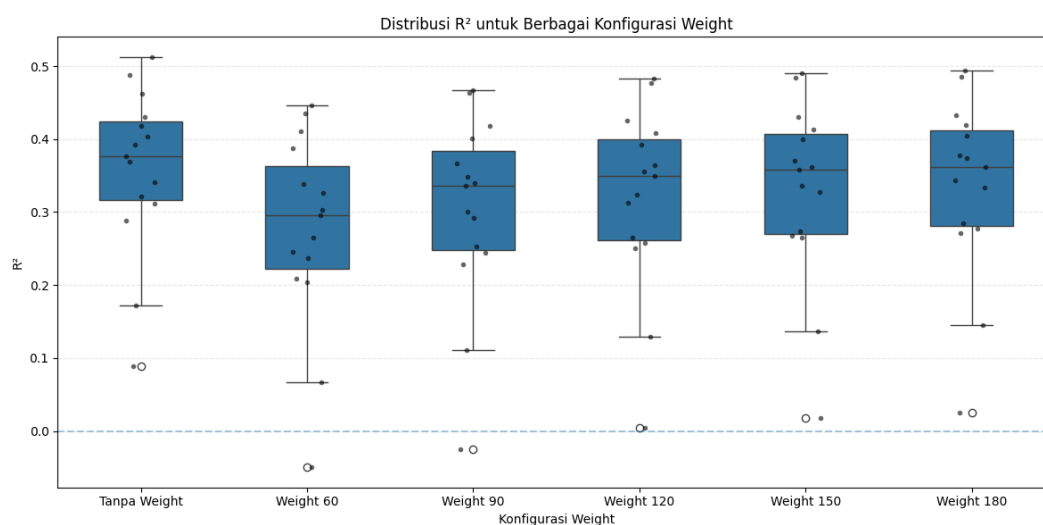
Tabel 4.6 Ringkasan Konfigurasi Terbaik pada Eksperimen 2: Perbandingan SVR dengan dan tanpa Pembobotan Waktu

<b>Wilayah</b>	<b>Konfigurasi Terbaik</b>	<b><math>R^2</math> Mean</b>	<b>MAE Mean</b>	<b>RMSE Mean</b>
Bengkulu	Pembobotan ( <i>Half-life 180</i> )	0.4330	19.9814	27.5653
Pekanbaru	Tanpa Pembobotan	0.3770	18.9871	26.3508
Sleman	Tanpa Pembobotan	0.3697	17.1405	23.1198
Malang	Tanpa Pembobotan	0.4033	16.7581	23.6372
Medan	Tanpa Pembobotan	0.5120	21.2028	28.7640
Kemayoran	Tanpa Pembobotan	0.4626	16.5324	21.7465
Banjarbaru	Tanpa Pembobotan	0.3923	12.9842	17.9793
Samarinda	Tanpa Pembobotan	0.2883	12.7232	16.7975
Jambi	Tanpa Pembobotan	0.3213	18.2876	25.8464
Semarang	Tanpa Pembobotan	0.4183	15.2253	20.0996
Sorong	Tanpa Pembobotan	0.0887	8.6948	12.9730
Kototabang	Tanpa Pembobotan	0.4876	8.9113	12.3306
Talang Betutu	Tanpa Pembobotan	0.3119	16.4952	22.9591
Pangkalanbun	Tanpa Pembobotan	0.3410	10.5683	14.4178
Maros	Tanpa Pembobotan	0.1721	13.1802	19.5320

Berdasarkan Tabel 4.6, penerapan pembobotan waktu eksponensial pada model *Support Vector Regression* (SVR) secara umum belum memberikan peningkatan performa dibandingkan model tanpa pembobotan. Pada sebagian besar wilayah penelitian, model tanpa pembobotan menghasilkan nilai  $R^2$  yang lebih tinggi dibandingkan seluruh konfigurasi pembobotan waktu eksponensial. Dari 15 wilayah penelitian, hanya wilayah Bengkulu yang menunjukkan sedikit peningkatan performa ketika menggunakan pembobotan dengan *half-life* 180 hari,

dengan nilai  $R^2$  sebesar 0.4330, sedikit lebih tinggi dibandingkan model tanpa pembobotan yang memiliki nilai  $R^2$  sebesar 0.4300. Sementara itu, pada wilayah selain Bengkulu, model tanpa pembobotan tetap memberikan nilai  $R^2$  yang lebih tinggi dibandingkan seluruh konfigurasi pembobotan. Hasil ini menunjukkan bahwa pemberian bobot yang lebih besar pada data terbaru tidak selalu meningkatkan kemampuan model dalam menjelaskan variasi konsentrasi PM<sub>2.5</sub>.

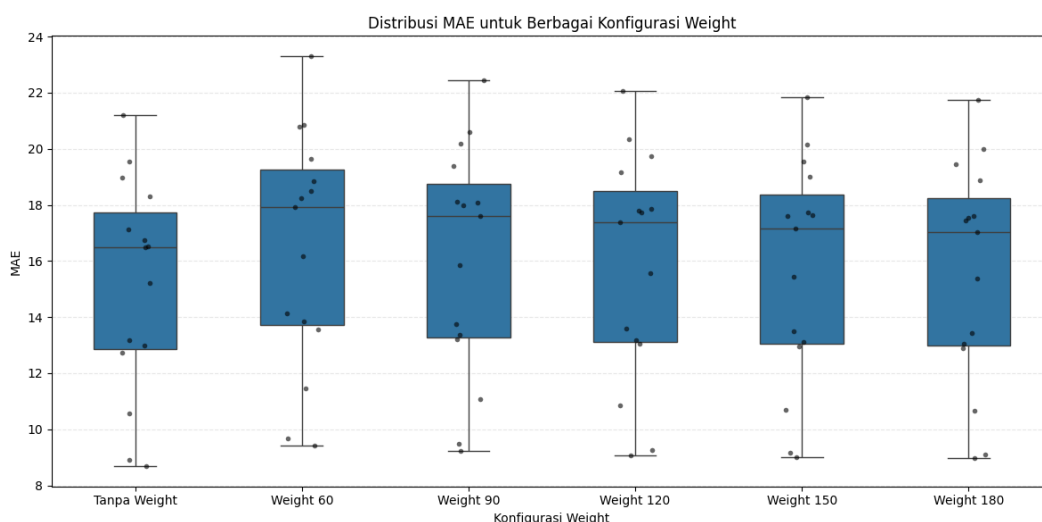
Untuk memperjelas perbandingan kemampuan model dalam menjelaskan variasi data, visualisasi nilai  $R^2$  untuk kedua skenario ditampilkan pada Gambar 4.9. Grafik ini menunjukkan bahwa hampir seluruh wilayah memiliki nilai  $R^2$  yang lebih tinggi pada model tanpa pembobotan dibandingkan model dengan pembobotan waktu eksponensial.



Gambar 4.9 Perbandingan Distribusi Nilai  $R^2$  Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial

Berdasarkan Gambar 4.9, distribusi nilai  $R^2$  pada berbagai konfigurasi pembobotan waktu eksponensial menunjukkan pola yang relatif konsisten di seluruh wilayah penelitian. Secara umum, nilai  $R^2$  model tanpa pembobotan cenderung lebih tinggi dibandingkan model dengan pembobotan waktu

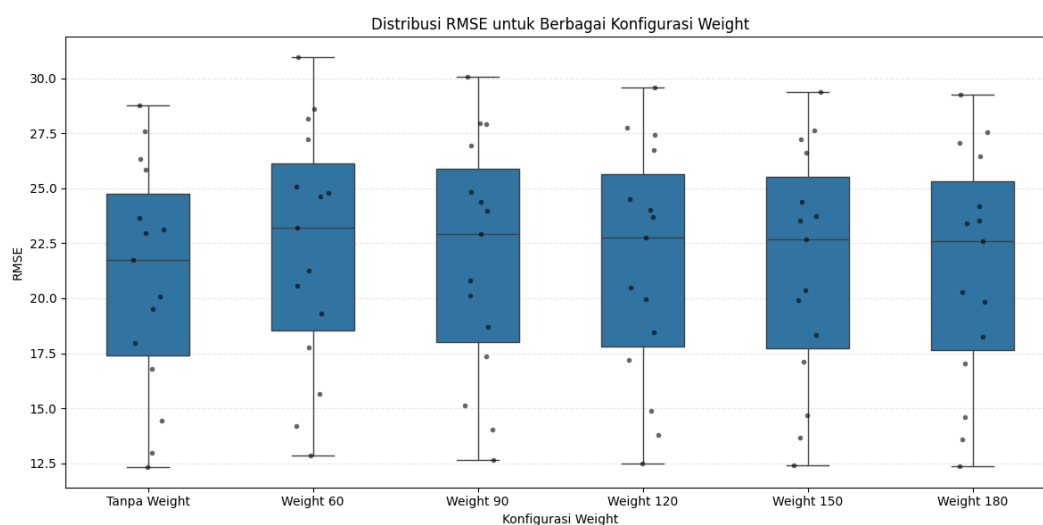
eksponensial. Meskipun nilai  $R^2$  pada model berbobot meningkat secara bertahap seiring bertambahnya nilai *half-life* dari 60 hingga 180 hari, peningkatan tersebut umumnya masih berada di bawah performa model tanpa pembobotan. Hal ini terlihat pada sebagian besar wilayah penelitian, di mana konfigurasi pembobotan tidak mampu melampaui nilai  $R^2$  model tanpa pembobotan. Kondisi tersebut menunjukkan bahwa meskipun pembobotan waktu eksponensial memberikan penekanan pada data terbaru, pendekatan tersebut belum mampu meningkatkan performa model secara signifikan dalam menjelaskan variasi data. Dengan demikian, penggunaan pembobotan waktu eksponensial tidak memberikan keuntungan yang konsisten terhadap performa model dalam hal kemampuan penjelasan variasi data. Untuk menganalisis perbandingan kesalahan absolut peramalan, distribusi nilai MAE pada kedua skenario ditampilkan pada Gambar 4.10.



Gambar 4.10 Perbandingan Distribusi Nilai MAE Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial

Berdasarkan Gambar 4.10, distribusi nilai MAE pada berbagai konfigurasi pembobotan waktu eksponensial menunjukkan bahwa model tanpa pembobotan

umumnya menghasilkan nilai kesalahan yang lebih kecil dibandingkan model berbobot. Pada sebagian besar wilayah penelitian, penerapan pembobotan waktu eksponensial dengan berbagai nilai *half-life* menghasilkan nilai MAE yang sedikit lebih tinggi dibandingkan konfigurasi tanpa pembobotan. Hal ini mengindikasikan bahwa kesalahan absolut peramalan cenderung meningkat ketika bobot yang lebih besar diberikan pada data yang lebih baru. Meskipun terdapat kecenderungan penurunan nilai MAE pada konfigurasi *half-life* yang lebih besar, khususnya pada *half-life* 150 hingga 180 hari, nilai MAE tersebut masih belum mampu mengungguli performa model tanpa pembobotan. Kondisi ini menunjukkan bahwa penekanan pada data terbaru tidak selalu meningkatkan akurasi peramalan model. Oleh karena itu, penerapan pembobotan waktu eksponensial perlu dipertimbangkan secara hati-hati karena dampaknya terhadap pengurangan kesalahan peramalan tidak terlihat secara konsisten pada seluruh wilayah penelitian. Selanjutnya, untuk mengevaluasi kesalahan peramalan yang lebih sensitif terhadap nilai ekstrem, distribusi nilai RMSE ditampilkan pada Gambar 4.11.



Gambar 4.11 Perbandingan Distribusi Nilai RMSE Model SVR dengan dan tanpa Pembobotan Waktu Eksponensial

Berdasarkan Gambar 4.11, distribusi nilai RMSE menunjukkan pola yang serupa dengan hasil pada MAE. Pada sebagian besar wilayah penelitian, konfigurasi dengan pembobotan waktu eksponensial menghasilkan nilai RMSE yang sedikit lebih tinggi dibandingkan model tanpa pembobotan. Hal ini mengindikasikan bahwa kesalahan peramalan kuadrat rata-rata cenderung meningkat ketika pembobotan waktu eksponensial diterapkan. Meskipun perbedaannya tidak selalu besar pada setiap wilayah, hasil ini mengindikasikan bahwa penggunaan pembobotan waktu eksponensial belum mampu memberikan peningkatan performa model secara signifikan.

Berdasarkan hasil evaluasi menggunakan metrik  $R^2$ , MAE, dan RMSE, penerapan pembobotan waktu eksponensial pada model *Support Vector Regression* (SVR) secara umum tidak memberikan peningkatan performa dibandingkan model tanpa pembobotan. Pada 14 dari 15 wilayah penelitian, model tanpa pembobotan menghasilkan nilai  $R^2$  yang lebih tinggi serta nilai MAE dan RMSE yang lebih rendah dibandingkan seluruh konfigurasi model berbobot. Meskipun performa model berbobot cenderung meningkat seiring bertambahnya nilai *half-life* hingga 180 hari, peningkatan tersebut masih belum mampu melampaui kinerja model tanpa pembobotan pada sebagian besar wilayah. Oleh karena itu, model SVR tanpa pembobotan waktu eksponensial dipilih sebagai konfigurasi yang lebih optimal untuk digunakan pada tahap pemodelan selanjutnya. Hasil ini menunjukkan bahwa penambahan kompleksitas melalui pembobotan waktu eksponensial tidak selalu



diperlukan, terutama ketika model dasar telah mampu menangkap pola data secara efektif.

#### 4.6.3.3. Eksperimen 3: Pengaruh Seleksi Fitur

Eksperimen ketiga dilakukan untuk menganalisis pengaruh seleksi fitur terhadap performa model *Support Vector Regression* (SVR) dalam peramalan konsentrasi PM2.5 satu hari ke depan. Seleksi fitur bertujuan untuk mengidentifikasi fitur-fitur yang paling relevan sehingga model dapat bekerja secara lebih efisien, mengurangi kompleksitas, serta berpotensi meningkatkan kemampuan generalisasi. Dalam konteks data runtun waktu kualitas udara yang memiliki banyak variabel turunan, proses seleksi fitur menjadi penting untuk meminimalkan pengaruh *noise* dan redundansi antar fitur.

Pada penelitian ini, proses rekayasa fitur menghasilkan total 26 fitur numerik yang mencakup variabel meteorologi, fitur *lag* PM2.5, *rolling window*, *first difference* serta *Exponential Weighted Moving Average* (EWMA). Namun demikian, tidak seluruh fitur tersebut selalu memberikan kontribusi positif terhadap pembelajaran model. Oleh karena itu, dilakukan pengujian seleksi fitur menggunakan metode *SelectKBest*, yang memilih sejumlah fitur terbaik berdasarkan skor statistik. Eksperimen ini membandingkan tiga konfigurasi utama, yaitu model SVR tanpa seleksi fitur, SVR dengan *SelectKBest*  $k=5$ , dan SVR dengan *SelectKBest*  $k=8$ . Setiap konfigurasi dilakukan *tuning hyperparameter* secara terpisah agar perbandingan performa yang dihasilkan bersifat adil. Ringkasan hasil eksperimen yang menampilkan konfigurasi terbaik pada setiap wilayah beserta nilai rata-rata  $R^2$ , MAE, dan RMSE disajikan pada Tabel 4.7. Tabel

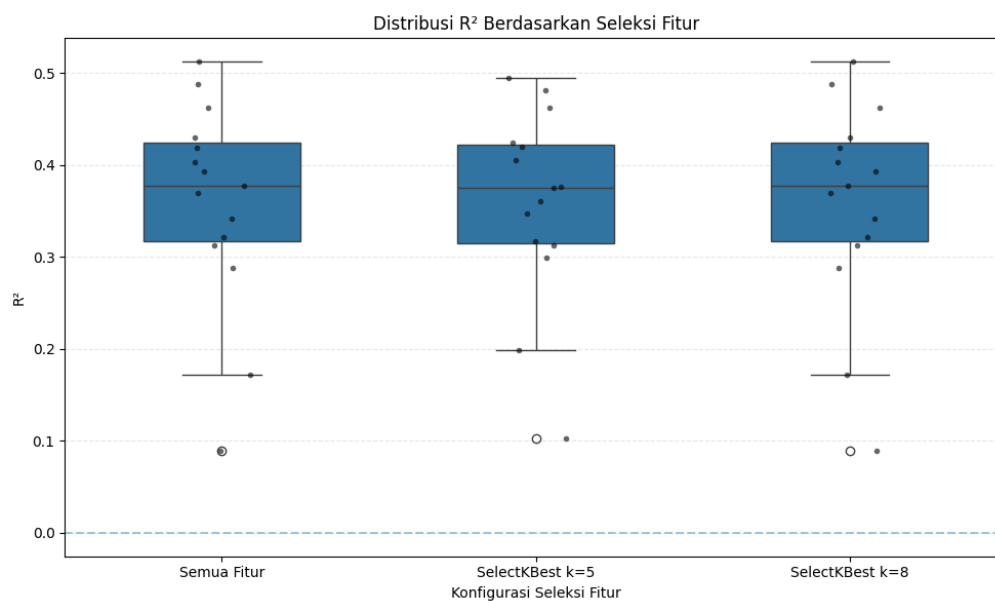
ini hanya menampilkan konfigurasi terbaik pada setiap wilayah, sedangkan perbandingan lengkap antar konfigurasi disajikan dalam bentuk visualisasi pada Gambar 4.12–4.14.

Tabel 4.7 Ringkasan Konfigurasi Terbaik pada Eksperimen 3: Perbandingan Tanpa Seleksi Fitur dan *SelectKBest*

<b>Wilayah</b>	<b>Konfigurasi Terbaik</b>	<b>R<sup>2</sup> Mean</b>	<b>MAE Mean</b>	<b>RMSE Mean</b>
Bengkulu	Semua Fitur / <i>SelectKBest</i> k=8	0.4300	19.5540	27.5801
Pekanbaru	Semua Fitur / <i>SelectKBest</i> k=8	0.3770	18.9871	26.3508
Sleman	<i>SelectKBest</i> k=5	0.3755	16.9624	22.9986
Malang	<i>SelectKBest</i> k=5	0.4196	16.4956	23.3168
Medan	Semua Fitur / <i>SelectKBest</i> k=8	0.5120	21.2028	28.7640
Kemayoran	Semua Fitur / <i>SelectKBest</i> k=8	0.4626	16.5324	21.7465
Banjarbaru	Semua Fitur / <i>SelectKBest</i> k=8	0.3923	12.9842	17.9793
Samarinda	<i>SelectKBest</i> k=5	0.3164	12.4040	16.5028
Jambi	Semua Fitur / <i>SelectKBest</i> k=8	0.3213	18.2876	25.8464
Semarang	<i>SelectKBest</i> k=5	0.4237	15.2585	20.0721
Sorong	<i>SelectKBest</i> k=5	0.1030	8.5998	12.8363
Kototabang	<i>SelectKBest</i> k=5	0.4943	8.8455	12.2466
Talang Betutu	Semua Fitur / <i>SelectKBest</i> k=8	0.3119	16.4952	22.9591
Pangkalanbun	<i>SelectKBest</i> k=5	0.3607	10.3701	14.2894
Maros	<i>SelectKBest</i> k=5	0.1988	12.8328	19.1788

Berdasarkan hasil pada Tabel 4.7, dapat diamati bahwa pengaruh seleksi fitur terhadap performa model bervariasi antarwilayah. Secara umum, konfigurasi *SelectKBest* k=8 menghasilkan nilai R<sup>2</sup>, MAE, dan RMSE yang identik dengan penggunaan semua fitur pada seluruh wilayah penelitian, sehingga jumlah fitur dapat dikurangi tanpa menurunkan kualitas peramalan. Sementara itu, konfigurasi *SelectKBest* k=5 menunjukkan hasil yang berbeda-beda antarwilayah. Peningkatan

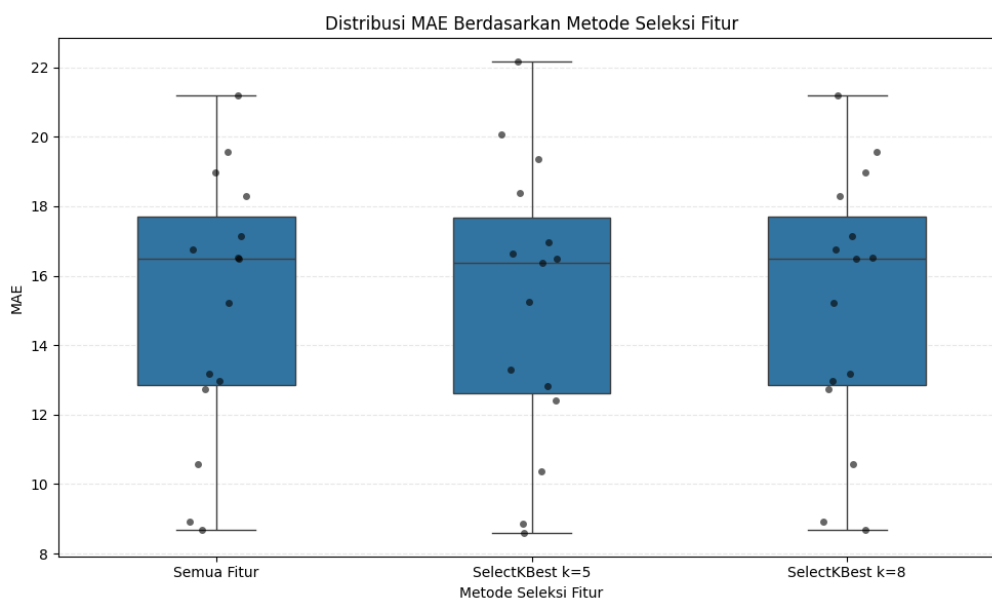
nilai  $R^2$  terjadi pada beberapa wilayah, yaitu Sleman, Malang, Samarinda, Semarang, Sorong, Kototabang, Pangkalanbun, dan Maros, sedangkan pada wilayah Bengkulu, Pekanbaru, Medan, Banjarbaru, Jambi, dan Talang Betutu penggunaan  $k=5$  justru menurunkan nilai  $R^2$ . Kondisi ini menunjukkan bahwa pengurangan fitur yang lebih agresif dapat membantu model pada wilayah tertentu, tetapi tidak memberikan keuntungan yang konsisten pada seluruh wilayah penelitian. Untuk memperjelas perbandingan performa antar konfigurasi, hasil evaluasi divisualisasikan dalam bentuk grafik. Perbandingan nilai  $R^2$  untuk ketiga konfigurasi seleksi fitur ditunjukkan pada Gambar 4.12.



Gambar 4.12 Perbandingan Distribusi Nilai  $R^2$  Model SVR Tanpa Seleksi Fitur, *SelectKBest*  $k=5$ , dan  $k=8$

Berdasarkan Gambar 4.12, distribusi nilai  $R^2$  untuk ketiga konfigurasi menunjukkan bahwa Semua Fitur dan *SelectKBest*  $k=8$  memiliki distribusi yang sama dan bertumpuk sepenuhnya. Hal ini menegaskan bahwa penggunaan delapan fitur terbaik sudah cukup untuk mempertahankan kemampuan model dalam

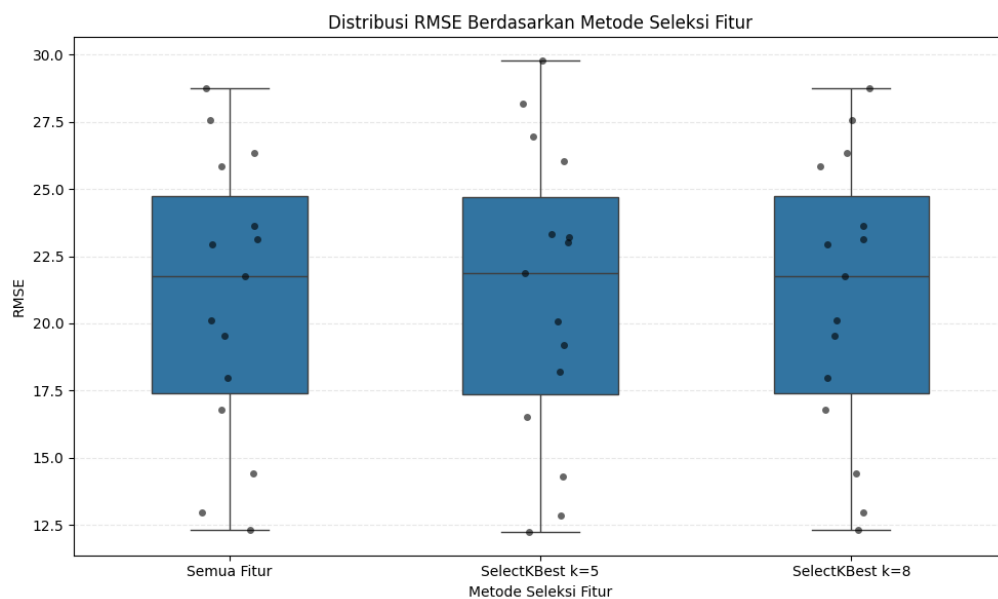
menjelaskan variasi data PM2.5. Sementara itu, distribusi nilai  $R^2$  pada *SelectKBest*  $k=5$  terlihat sedikit berbeda. Pada beberapa wilayah, konfigurasi ini menghasilkan nilai  $R^2$  yang lebih tinggi, seperti pada Malang, Samarinda, Sorong, Kototabang, Pangkalanbun, dan Maros, namun pada wilayah lain justru lebih rendah. Dengan demikian, pengurangan jumlah fitur menjadi lima dapat memberikan manfaat pada kondisi tertentu, tetapi tidak menunjukkan peningkatan performa yang konsisten secara keseluruhan. Oleh karena itu, untuk memperoleh gambaran yang lebih lengkap mengenai performa model, analisis dilanjutkan pada metrik kesalahan absolut yang ditunjukkan pada Gambar 4.13.



Gambar 4.13 Perbandingan Distribusi Nilai MAE Model SVR Tanpa Seleksi Fitur, *SelectKBest*  $k=5$ , dan  $k=8$

Berdasarkan Gambar 4.13, distribusi nilai MAE menunjukkan bahwa konfigurasi Semua Fitur dan *SelectKBest*  $k=8$  kembali menghasilkan nilai yang identik pada seluruh wilayah. Hal ini menunjukkan bahwa reduksi fitur menjadi delapan tidak mengubah tingkat kesalahan absolut model. Pada konfigurasi

*SelectKBest* k=5, terlihat adanya perubahan nilai MAE yang relatif kecil antarwilayah. Di beberapa wilayah seperti Malang, Samarinda, Sorong, Kototabang, Pangkalanbun, dan Maros, nilai MAE menjadi lebih kecil, yang menandakan bahwa model mampu menghasilkan peramalan yang sedikit lebih akurat. Namun pada beberapa wilayah lain, seperti Bengkulu, Pekanbaru, Medan, Banjarbaru, Jambi, dan Talang Betutu, nilai MAE justru sedikit meningkat. Hal ini menunjukkan bahwa penggunaan lima fitur terbaik dapat membantu menurunkan kesalahan peramalan pada wilayah tertentu, tetapi pengaruhnya belum menunjukkan keunggulan yang konsisten secara keseluruhan. Selanjutnya, perbandingan nilai RMSE pada tiap konfigurasi ditunjukkan pada Gambar 4.14.



Gambar 4.14 Perbandingan Distribusi Nilai RMSE Model SVR Tanpa Seleksi Fitur, *SelectKBest* k=5, dan k=8

Berdasarkan Gambar 4.14, distribusi nilai RMSE menunjukkan pola yang serupa dengan MAE. Konfigurasi Semua Fitur dan *SelectKBest* k=8 menghasilkan nilai RMSE yang sama pada seluruh wilayah, sehingga dapat disimpulkan bahwa

penggunaan delapan fitur terbaik tidak mempengaruhi besar kecilnya kesalahan kuadrat rata-rata. Sementara itu, konfigurasi *SelectKBest*  $k=5$  menunjukkan sedikit perubahan pada beberapa wilayah. Pada wilayah seperti Malang, Samarinda, Sorong, Kototabang, Pangkalanbun, dan Maros, nilai RMSE cenderung lebih rendah, sedangkan pada wilayah lain justru sedikit lebih tinggi. Pola ini menunjukkan bahwa pengurangan fitur menjadi lima dapat mengurangi *noise* dan memperbaiki generalisasi model pada wilayah tertentu, tetapi efeknya tidak merata. Secara umum, perbedaan nilai RMSE antar konfigurasi masih relatif kecil.

Secara keseluruhan, penerapan seleksi fitur pada Eksperimen 3 belum memberikan peningkatan performa yang signifikan terhadap model secara umum. Konfigurasi *SelectKBest* dengan  $k=8$  mampu mempertahankan performa yang setara dengan penggunaan seluruh fitur, sehingga dapat dianggap sebagai bentuk penyederhanaan model tanpa kehilangan akurasi. Hal ini menunjukkan bahwa sebagian besar informasi penting telah direpresentasikan oleh delapan fitur terpilih, sementara fitur lainnya memiliki kontribusi yang relatif kecil. Sementara itu, konfigurasi *SelectKBest*  $k=5$  menunjukkan hasil yang bervariasi antarwilayah, dengan peningkatan performa pada beberapa wilayah namun penurunan pada wilayah lainnya. Oleh karena itu, konfigurasi *SelectKBest*  $k=8$  dipilih sebagai pendekatan yang lebih stabil dan efisien untuk tahap pemodelan selanjutnya karena mampu menjaga keseimbangan antara akurasi dan kompleksitas model.

#### **4.6.3.4. Eksperimen 4: Pengaruh Transformasi Target**

Eksperimen keempat dilakukan untuk mengevaluasi pengaruh transformasi target terhadap performa model *Support Vector Regression* (SVR) dalam

meramalkan konsentrasi PM2.5. Transformasi target merupakan salah satu teknik yang umum digunakan pada pemodelan regresi, terutama ketika distribusi data target bersifat tidak normal, memiliki *skewness* tinggi, atau mengandung nilai ekstrem.

Pada data kualitas udara, konsentrasi PM2.5 sering kali menunjukkan fluktuasi yang tajam akibat perubahan cuaca, aktivitas manusia, maupun kejadian polusi tertentu. Kondisi ini menyebabkan distribusi nilai PM2.5 cenderung tidak simetris dan sulit dipelajari secara optimal oleh model SVR dalam skala asli. Untuk mengatasi hal tersebut, eksperimen ini membandingkan dua skenario, yaitu model SVR tanpa transformasi target dan model SVR dengan transformasi target menggunakan metode *Yeo-Johnson*. Metode ini dipilih karena mampu menangani data bernilai nol maupun negatif, serta sering digunakan untuk menstabilkan variansi data pada regresi *non-linear*. Kedua skenario dilakukan dengan proses *tuning hyperparameter* agar perbandingan performa tetap adil. Ringkasan hasil eksperimen yang menampilkan konfigurasi terbaik pada setiap wilayah beserta nilai rata-rata  $R^2$ , MAE, dan RMSE disajikan pada Tabel 4.8. Tabel ini hanya menampilkan konfigurasi terbaik pada setiap wilayah, sedangkan perbandingan lengkap antar konfigurasi disajikan dalam bentuk visualisasi pada Gambar 4.15–4.17.

Tabel 4.8 Ringkasan Konfigurasi Terbaik pada Eksperimen 4: Pengaruh Transformasi Target *Yeo-Johnson*

Wilayah	Konfigurasi Terbaik	$R^2$ Mean	MAE Mean	RMSE Mean
Bengkulu	<i>Yeo-Johnson</i>	0.4300	19.5540	27.5801
Pekanbaru	<i>Yeo-Johnson</i>	0.3770	18.9871	26.3508
Sleman	<i>Yeo-Johnson</i>	0.3697	17.1405	23.1198
Malang	<i>Yeo-Johnson</i>	0.4033	16.7581	23.6372

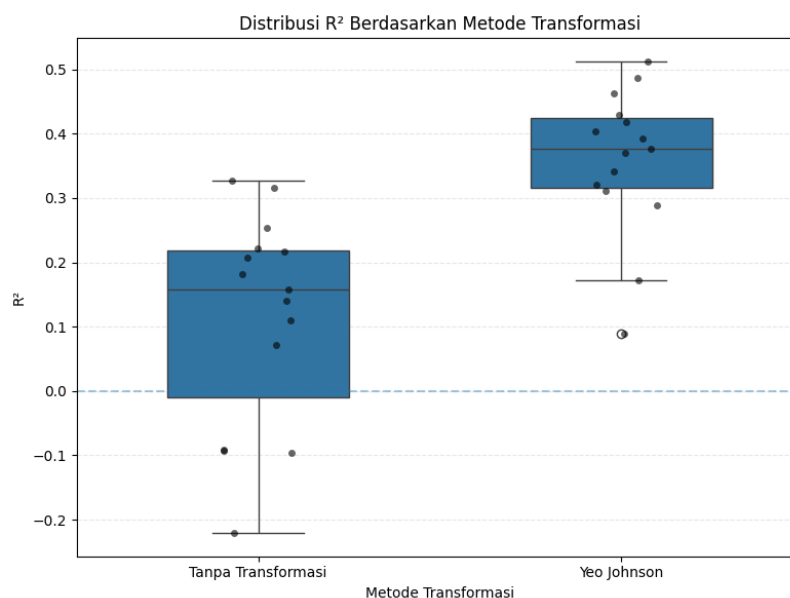
Medan	<i>Yeo-Johnson</i>	0.5120	21.2028	28.7640
Kemayoran	<i>Yeo-Johnson</i>	0.4626	16.5324	21.7465
Banjarbaru	<i>Yeo-Johnson</i>	0.3923	12.9842	17.9793
Samarinda	<i>Yeo-Johnson</i>	0.2883	12.7232	16.7975
Jambi	<i>Yeo-Johnson</i>	0.3213	18.2876	25.8464
Semarang	<i>Yeo-Johnson</i>	0.4183	15.2253	20.0996
Sorong	<i>Yeo-Johnson</i>	0.0887	8.6948	12.9730
Kototabang	<i>Yeo-Johnson</i>	0.4876	8.9113	12.3306
Talang Betutu	<i>Yeo-Johnson</i>	0.3119	16.4952	22.9591
Pangkalanbun	<i>Yeo-Johnson</i>	0.3410	10.5683	14.4178
Maros	<i>Yeo-Johnson</i>	0.1721	13.1802	19.5320

Tabel 4.8 menyajikan hasil eksperimen mengenai pengaruh transformasi target menggunakan metode *Yeo-Johnson* terhadap performa model *Support Vector Regression*. Berdasarkan hasil evaluasi dan distribusi performa yang ditunjukkan pada grafik boxplot, model dengan transformasi *Yeo-Johnson* menunjukkan performa yang lebih stabil dibandingkan model tanpa transformasi target. Nilai  $R^2$  tertinggi diperoleh pada wilayah Medan (0.5120), diikuti oleh Kototabang (0.4876) dan Kemayoran (0.4626). Sementara itu, wilayah Sorong (0.0887) dan Maros (0.1721) memiliki nilai  $R^2$  yang relatif lebih rendah dibandingkan wilayah lainnya. Meskipun demikian, nilai MAE pada wilayah tersebut tetap relatif kecil, yaitu 8.6948 pada Sorong dan 13.1802 pada Maros. Secara umum, hasil pada tabel menunjukkan bahwa transformasi *Yeo-Johnson* mampu meningkatkan kemampuan model dalam menjelaskan variasi data sekaligus menjaga tingkat kesalahan peramalan pada nilai yang relatif rendah. Selain itu, beberapa wilayah seperti Kemayoran, Semarang, Sorong, dan Maros menunjukkan nilai  $R^2$  negatif sebelum transformasi, yang mengindikasikan bahwa model belum mampu menjelaskan variasi data dengan baik. Setelah penerapan transformasi *Yeo-Johnson*, seluruh nilai



$R^2$  menjadi positif, yang menunjukkan peningkatan kemampuan model secara signifikan.

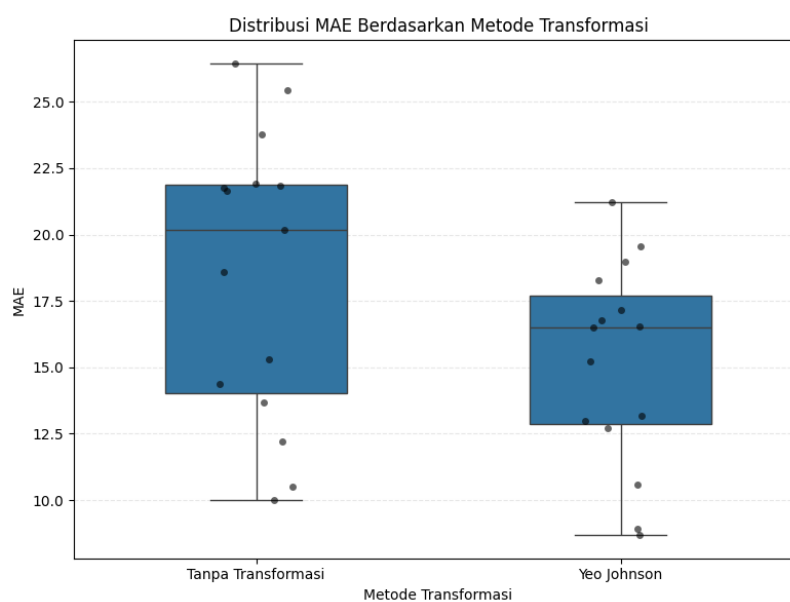
Perbandingan nilai  $R^2$  tersebut divisualisasikan lebih lanjut pada Gambar 4.15. Grafik ini memperlihatkan secara jelas bahwa transformasi *Yeo-Johnson* mampu meningkatkan nilai  $R^2$  pada seluruh wilayah penelitian. Tidak ditemukan wilayah yang mengalami penurunan nilai  $R^2$  setelah transformasi diterapkan, yang menunjukkan konsistensi dampak positif dari pendekatan ini terhadap kemampuan model dalam menangkap pola data PM2.5.



Gambar 4.15 Perbandingan Distribusi Nilai  $R^2$  Model SVR Tanpa Transformasi Target dan dengan Transformasi *Yeo-Johnson*

Berdasarkan Gambar 4.15, distribusi nilai  $R^2$  antara model tanpa transformasi target dan model dengan transformasi *Yeo-Johnson* menunjukkan perbedaan yang jelas. Model tanpa transformasi target memiliki variasi nilai  $R^2$  yang lebih besar antarwilayah, sedangkan penggunaan transformasi *Yeo-Johnson* menghasilkan nilai  $R^2$  yang lebih tinggi dan lebih stabil. Hal ini menunjukkan

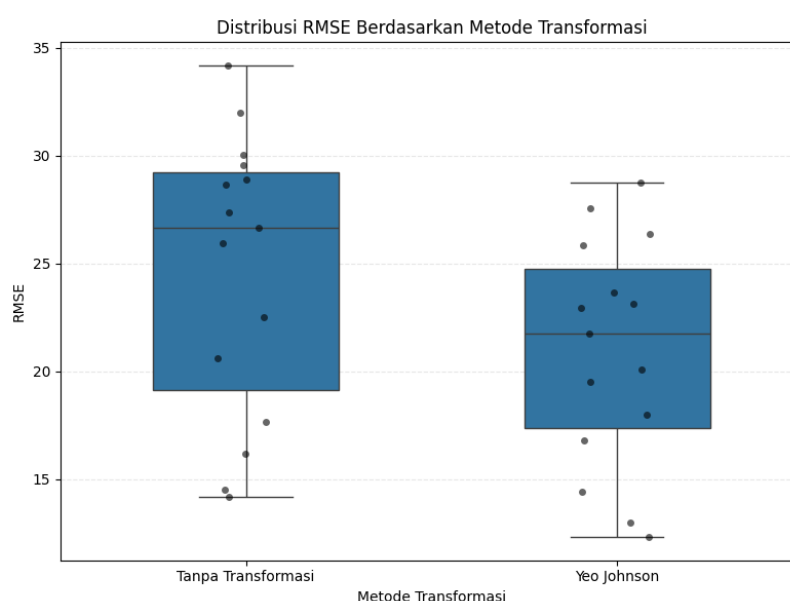
bahwa transformasi target membantu memperbaiki distribusi data sehingga hubungan antarvariabel lebih mudah dipelajari oleh model. Dengan distribusi data yang lebih stabil, model SVR mampu menangkap pola variasi PM2.5 dengan lebih baik. Untuk memperoleh gambaran yang lebih lengkap mengenai kesalahan peramalan, selanjutnya akan dibahas perbandingan nilai *Mean Absolute Error* (MAE) pada Gambar 4.16.



Gambar 4.16 Perbandingan Distribusi Nilai MAE Model SVR Tanpa Transformasi Target dan dengan Transformasi *Yeo-Johnson*

Berdasarkan Gambar 4.16, distribusi nilai MAE menunjukkan bahwa penerapan transformasi *Yeo-Johnson* menghasilkan nilai kesalahan absolut yang lebih rendah dibandingkan model tanpa transformasi target. Pada sebagian besar wilayah, nilai MAE setelah transformasi terlihat lebih kecil dan lebih terkonsentrasi pada rentang yang sempit. Hal ini menunjukkan bahwa model mampu menghasilkan peramalan yang lebih dekat dengan nilai aktual ketika transformasi target diterapkan. Sebaliknya, model tanpa transformasi target menunjukkan variasi

nilai MAE yang lebih besar dan cenderung menghasilkan kesalahan yang lebih tinggi. Dengan demikian, transformasi *Yeo-Johnson* tidak hanya meningkatkan kemampuan model dalam menjelaskan variasi data, tetapi juga membantu mengurangi kesalahan peramalan secara langsung. Selanjutnya, untuk mengevaluasi kesalahan peramalan yang lebih sensitif terhadap nilai ekstrem, perbandingan *Root Mean Squared Error* (RMSE) ditunjukkan pada Gambar 4.17.



Gambar 4.17 Perbandingan Distribusi Nilai RMSE Model SVR Tanpa Transformasi Target dan dengan Transformasi *Yeo-Johnson*

Berdasarkan Gambar 4.17, distribusi nilai RMSE menunjukkan pola yang serupa dengan MAE. Model dengan transformasi *Yeo-Johnson* menghasilkan nilai RMSE yang lebih rendah dan lebih konsisten dibandingkan model tanpa transformasi target. Nilai RMSE yang lebih kecil menunjukkan bahwa kesalahan peramalan yang lebih besar dapat diminimalkan dengan lebih baik ketika transformasi diterapkan. Sebaliknya, pada model tanpa transformasi target, nilai RMSE cenderung lebih tinggi dan lebih bervariasi antarwilayah. Hal ini

menunjukkan bahwa tanpa transformasi, model lebih rentan terhadap pengaruh nilai ekstrem dalam data. Dengan demikian, transformasi *Yeo-Johnson* membantu meningkatkan stabilitas peramalan model.

Secara keseluruhan, penerapan transformasi *Yeo-Johnson* memberikan peningkatan performa model yang konsisten pada seluruh wilayah penelitian. Transformasi ini meningkatkan nilai  $R^2$ , sekaligus menurunkan nilai MAE dan RMSE dibandingkan model tanpa transformasi target. Hasil ini menunjukkan bahwa transformasi target mampu memperbaiki distribusi data PM2.5 sehingga hubungan antarvariabel menjadi lebih mudah dipelajari oleh model. Dengan distribusi data yang lebih stabil, model SVR dapat menghasilkan peramalan yang lebih akurat dan lebih konsisten. Oleh karena itu, transformasi *Yeo-Johnson* dipilih sebagai pendekatan transformasi target yang digunakan pada tahap pemodelan selanjutnya dalam penelitian ini.

#### **4.7 Pembahasan Hasil**

Pada subbab ini membahas hasil evaluasi model *Support Vector Regression* (SVR) dalam meramalkan konsentrasi PM2.5 satu hari ke depan (H+1) di seluruh wilayah penelitian. Analisis difokuskan pada interpretasi performa model berdasarkan metrik evaluasi, perbandingan dengan *baseline*, variasi performa antar-wilayah, serta faktor-faktor yang memengaruhi perbedaan performa. Diskusi ini juga mengintegrasikan temuan dari empat eksperimen terbatas yang dilakukan untuk mengeksplorasi sensitivitas model terhadap konfigurasi fitur, pembobotan waktu, seleksi fitur, dan transformasi target.

Hasil evaluasi menunjukkan bahwa model SVR secara konsisten mampu menangkap sebagian besar pola temporal PM2.5, meskipun performa antar-wilayah bervariasi. Koefisien determinasi ( $R^2$ ) untuk model SVR berada pada kisaran 0.089 hingga 0.512, dengan nilai tertinggi di Medan (0.512), diikuti Kototabang (0.488) dan Kemayoran (0.463). Wilayah dengan  $R^2$  rendah, seperti Sorong (0.089) dan Maros (0.172), menunjukkan bahwa model kesulitan menjelaskan variasi data di wilayah tersebut, meskipun MAE dan RMSE relatif rendah, menunjukkan bahwa fluktuasi harian mungkin lebih kecil tetapi pola variasi tidak kompleks sehingga model tidak mendapatkan informasi yang cukup untuk  $R^2$  tinggi. MAE berkisar antara 8.69 hingga 21.20  $\mu\text{g}/\text{m}^3$ , sedangkan RMSE antara 12.33 hingga 28.76  $\mu\text{g}/\text{m}^3$ . Perbedaan ini mencerminkan tingkat fluktuasi lokal dan tingkat *noise* dalam data PM2.5 dan variabel meteorologi, termasuk pengaruh musiman, aktivitas manusia, dan kondisi atmosfer setempat.

Perbandingan dengan model *baseline* menegaskan keunggulan SVR. Model *baseline* yang menggunakan nilai PM2.5 periode sebelumnya sebagai peramalan menunjukkan  $R^2$  negatif pada hampir semua wilayah, bahkan mencapai  $-4.361$  pada Talang Betutu, menunjukkan ketidakmampuan menangkap pola temporal dan performa yang lebih buruk dibandingkan peramalan sederhana. MAE dan RMSE *baseline* jauh lebih tinggi, menunjukkan peramalan yang kurang akurat dan tidak memadai untuk digunakan sebagai acuan peramalan H+1. Model SVR, dengan  $R^2$  positif dan distribusi MAE/RMSE yang lebih stabil, membuktikan kemampuan model untuk menangkap pola historis dengan baik dan menghasilkan peramalan yang lebih dapat diandalkan.

Analisis eksperimen terbatas menambah pemahaman tentang sensitivitas model terhadap berbagai konfigurasi. Eksperimen pertama menyoroti pengaruh panjang *lag* dan *rolling window*. Hasil menunjukkan bahwa *Lag 1 Rolling Window 3* secara umum memberikan performa terbaik, dengan median  $R^2$  tinggi, MAE dan RMSE rendah di sebagian besar wilayah. *Lag* pendek lebih efektif dalam menangkap pola jangka pendek, mengurangi sensitivitas terhadap *outlier* dan fluktuasi ekstrem. *Lag* panjang, seperti *Lag 7 Rolling Window 14*, cenderung menurunkan  $R^2$  dan meningkatkan kesalahan peramalan karena model memasukkan data historis yang kurang relevan dan *noise* yang lebih banyak. Beberapa wilayah, seperti Pekanbaru dan Maros, menunjukkan performa optimal pada *Lag 3 Rolling Window 7*, yang mengindikasikan bahwa karakteristik data lokal mungkin memerlukan informasi historis jangka menengah untuk menangkap pola spesifik. *Boxplot* mendukung temuan ini dengan menunjukkan distribusi  $R^2$  yang lebih sempit dan *outlier* lebih sedikit pada konfigurasi terbaik, sedangkan *lag* panjang memiliki IQR lebih lebar dan *outlier* lebih tinggi.

Eksperimen kedua menguji pengaruh pembobotan waktu eksponensial. Model dengan pembobotan ini, secara umum, tidak menunjukkan peningkatan performa dibandingkan model tanpa pembobotan. Median  $R^2$  untuk model tanpa pembobotan lebih tinggi, MAE dan RMSE lebih rendah pada mayoritas wilayah, kecuali Bengkulu dengan *half-life* 180 hari yang menunjukkan sedikit peningkatan  $R^2$ . Distribusi *boxplot* menunjukkan bahwa wilayah dengan fluktuasi ekstrem atau *outlier* cenderung mengalami penurunan performa ketika bobot lebih besar

diberikan pada data terbaru, menunjukkan bahwa pemberian bobot tidak selalu menguntungkan jika pola historis relevan tersebar di seluruh periode pengamatan.

Seleksi fitur diuji melalui *SelectKBest* dengan  $k=8$  dan  $k=5$ . Konfigurasi  $k=8$  mempertahankan performa identik dengan penggunaan semua 26 fitur, menunjukkan bahwa delapan fitur terbaik cukup untuk mempertahankan akurasi. Konfigurasi  $k=5$  meningkatkan  $R^2$  dan menurunkan kesalahan pada beberapa wilayah seperti Malang, Samarinda, Sorong, Kototabang, Pangkalanbun, dan Maros, tetapi menurunkan performa wilayah lain. Hal ini menunjukkan bahwa pengurangan fitur yang agresif dapat meningkatkan generalisasi pada kondisi tertentu, tetapi tidak dapat diterapkan secara universal. *Boxplot* memperlihatkan bahwa distribusi  $R^2$  dan MAE lebih stabil untuk  $k=8$  dibanding  $k=5$ , dan *outlier* lebih sedikit pada konfigurasi optimal.

Transformasi target menggunakan *Yeo-Johnson* menunjukkan dampak positif yang paling konsisten. Model dengan transformasi memiliki  $R^2$  lebih tinggi, MAE dan RMSE lebih rendah, distribusi lebih stabil, IQR lebih sempit, dan *outlier* lebih sedikit. Wilayah dengan  $R^2$  rendah pada model tanpa transformasi, seperti Sorong dan Maros, menunjukkan peningkatan yang nyata setelah transformasi. Hal ini menegaskan bahwa sebagian perbedaan performa antar-wilayah disebabkan oleh distribusi data yang *skewed*, nilai ekstrem, dan fluktuasi yang tidak merata, yang dapat diatasi dengan transformasi target untuk menstabilkan data.

Faktor penyebab variasi performa antar-wilayah dapat diuraikan sebagai berikut. Pertama, distribusi data lokal yang stabil memungkinkan model menangkap pola temporal dengan lebih baik. Kedua, fluktuasi ekstrem dan *outlier*

meningkatkan MAE dan RMSE, serta menurunkan  $R^2$ , khususnya pada *lag* panjang atau model tanpa transformasi. Ketiga, *noise* dari variabel meteorologi, aktivitas manusia, dan kondisi atmosfer yang kompleks menambah kesulitan peramalan. Keempat, kombinasi panjang *lag*, *rolling window*, jumlah fitur, dan penerapan transformasi target mempengaruhi sensitivitas model terhadap karakteristik data lokal. Terakhir, transformasi target terbukti menstabilkan distribusi, mengurangi pengaruh *outlier*, dan meningkatkan konsistensi performa.

Secara keseluruhan, model SVR terbukti unggul dibandingkan *baseline* dalam meramalkan konsentrasi PM<sub>2.5</sub> H+1. Interpretasi hasil eksperimen menunjukkan bahwa konfigurasi fitur, transformasi target, dan pengaturan parameter model harus disesuaikan dengan karakteristik data lokal untuk mencapai performa optimal. Analisis ini memberikan dasar ilmiah yang kuat untuk pengembangan sistem prediksi kualitas udara selanjutnya yang adaptif terhadap kondisi setiap wilayah penelitian. Pemahaman faktor-faktor yang memengaruhi variasi performa antar-wilayah memungkinkan peramalan lebih akurat dan dapat dimanfaatkan untuk langkah antisipatif, seperti pengelolaan risiko kesehatan masyarakat dan pengambilan keputusan terkait polusi udara.

#### **4.8 Integrasi Islam**

Rendahnya nilai koefisien determinasi ( $R^2$ ) di sebagian wilayah menunjukkan bahwa variasi konsentrasi PM<sub>2.5</sub> belum sepenuhnya dapat dijelaskan oleh model, menandakan kompleksitas pencemaran udara yang dipengaruhi berbagai faktor seperti cuaca, transportasi, industri, dan dinamika atmosfer, serta keterbatasan data. Meski demikian, metode *Support Vector Regression* (SVR) tetap



mampu menangkap sebagian pola historis dan menghasilkan peramalan yang lebih baik dibandingkan pendekatan sederhana. Peramalan H+1 dapat dimanfaatkan sebagai informasi awal untuk langkah antisipatif, seperti membatasi aktivitas luar ruang dan meningkatkan kewaspadaan terhadap risiko kesehatan. Upaya pencegahan ini tidak hanya bernilai ilmiah, tetapi juga memiliki dimensi moral dalam Islam. Allah SWT berfirman dalam Surah Al-Baqarah ayat 205:

وَإِذَا تَوَلَّى سَعَىٰ فِي الْأَرْضِ لِيُفْسِدَ فِيهَا وَيُهْلِكَ أَحْرَثَ وَالنَّسْلَ ۗ وَاللَّهُ لَا يُحِبُّ الْفُسَادَ

*“Dan apabila ia berpaling (darimu), ia berjalan di bumi untuk mengadakan kerusakan padanya, dan merusak tanam-tanaman dan binatang ternak, dan Allah tidak menyukai kebinasaan” (QS. Al-Baqarah:205).*

Ayat ini menegaskan bahwa segala tindakan yang menimbulkan kerusakan di bumi hingga berdampak pada rusaknya sumber kehidupan manusia dan makhluk lain merupakan perbuatan yang tidak disukai oleh Allah SWT. Ulama tafsir seperti Ibnu Katsir menjelaskan bahwa kerusakan yang dimaksud merujuk pada perbuatan yang menimbulkan kerusakan di muka bumi, seperti merusak sumber kehidupan manusia berupa tanaman dan hewan, sehingga menghilangkan kemaslahatan serta mengganggu keberlangsungan kehidupan (Shaykh & Ghoffar, 2006). Dalam konteks modern, pencemaran udara akibat aktivitas manusia yang meningkatkan konsentrasi polutan seperti PM2.5 dapat dipahami sebagai salah satu bentuk kerusakan lingkungan, sehingga upaya pemantauan dan peramalan kualitas udara menjadi bagian dari ikhtiar untuk mencegah dampak tersebut serta menjaga keberlanjutan kehidupan. Selain itu, sebagai penguatan terhadap larangan melakukan kerusakan, Allah SWT juga berfirman dalam Surah Al-A'raf ayat 56:

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ حَوْفًا وَقَطْمَعًا ۗ إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ

*“Dan janganlah kamu membuat kerusakan di muka bumi, sesudah (Allah) memperbaikinya. Dan berdoalah kepada-Nya dengan rasa takut (tidak akan diterima) dan harapan (akan dikabulkan). Sesungguhnya rahmat Allah amat dekat kepada orang-orang yang berbuat baik.” (QS. Ar- A'raf:56).*

Ayat ini mengandung peringatan agar manusia tidak merusak keseimbangan alam yang telah diciptakan Allah dalam keadaan teratur dan harmonis. Ibnu Katsir menjelaskan bahwa larangan tersebut mencakup segala bentuk perbuatan yang menimbulkan kerusakan dan bahaya di muka bumi setelah Allah memperbaiki serta menata keadaannya dengan baik, karena kerusakan yang terjadi setelah adanya keteraturan akan lebih membahayakan kehidupan manusia (Shaykh & Ghoffar, 2006). Selain itu, menjaga kualitas udara berkaitan erat dengan upaya melindungi kehidupan manusia. Paparan PM2.5 dalam kadar tinggi dapat meningkatkan risiko gangguan pernapasan, penyakit jantung, dan berbagai masalah kesehatan lainnya. Dengan menyediakan informasi peramalan kualitas udara, sistem yang dikembangkan dalam penelitian ini berkontribusi dalam mendukung upaya perlindungan kesehatan masyarakat melalui pendekatan pencegahan.

Penelitian ini tidak hanya memiliki nilai ilmiah dalam pengembangan model peramalan berbasis *Support Vector Regression*, tetapi juga mencerminkan tanggung jawab etis terhadap lingkungan dan keselamatan kehidupan. Pengembangan sistem peramalan tidak semata berfokus pada akurasi, melainkan juga pada kontribusi nyata bagi kesejahteraan masyarakat. Integrasi ilmu pengetahuan dan nilai-nilai Islam menegaskan bahwa pemanfaatan teknologi merupakan bagian dari tanggung jawab manusia dalam mengelola bumi secara bijaksana dan berkelanjutan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

- a. Data kualitas udara yang diperoleh melalui *World Air Quality Index* (WAQI) API dapat dimanfaatkan untuk membangun dataset peramalan setelah melalui tahapan prapemrosesan yang meliputi pengurutan kronologis, pembersihan data, serta rekayasa fitur. Dengan proses tersebut, data yang semula tidak terstruktur dapat diubah menjadi dataset yang siap digunakan dalam pemodelan peramalan konsentrasi PM<sub>2.5</sub>.
- b. Metode *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF) dapat digunakan untuk membangun model peramalan konsentrasi PM<sub>2.5</sub> dan mampu menghasilkan performa peramalan yang baik, yang ditunjukkan melalui nilai evaluasi seperti MAE, RMSE, dan koefisien determinasi ( $R^2$ ).
- c. *Pipeline* sistem yang mencakup proses pengambilan data dari WAQI API, praproses data, pelatihan model, hingga penyajian hasil peramalan berhasil dirancang dan diimplementasikan dalam bentuk sistem informasi berbasis web yang dapat digunakan untuk memantau kualitas udara pada 15 wilayah di Indonesia.

## 5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, beberapa saran untuk pengembangan penelitian dan sistem di masa mendatang adalah sebagai berikut:

- a. Penelitian selanjutnya disarankan menggunakan data historis dengan jumlah yang lebih besar, seperti data per jam, agar model dapat mempelajari fluktuasi PM2.5 secara lebih detail dan meningkatkan akurasi peramalan.
- b. Penambahan variabel eksternal seperti curah hujan, radiasi matahari, arah angin, kepadatan lalu lintas, aktivitas industri, serta data kebakaran hutan dan lahan dapat membantu model menangkap faktor yang memengaruhi perubahan konsentrasi PM2.5 secara lebih komprehensif.
- c. Penelitian selanjutnya dapat membandingkan metode *Support Vector Regression* dengan metode lain seperti *Long Short-Term Memory* (LSTM), model berbasis *gradient boosting* seperti XGBoost dan LightGBM untuk mengetahui metode yang paling optimal dalam peramalan kualitas udara.
- d. Pada sisi sistem, aplikasi web dapat dikembangkan dengan penambahan fitur dashboard analitik, visualisasi tren historis, notifikasi peringatan dini berbasis ambang batas kesehatan.

Dengan pengembangan lebih lanjut berdasarkan saran tersebut, sistem peramalan kualitas udara yang telah dikembangkan diharapkan dapat memberikan manfaat yang lebih luas dalam konteks akademik, sosial, dan lingkungan, serta mendukung upaya mitigasi risiko kesehatan akibat polusi udara secara berkelanjutan

## DAFTAR PUSTAKA

- Abdelmalek, M. M., Mahmoud, H., & Shokry, H. (2025). Prognosis of Air Quality Index and Air Pollution Using Machine Learning Techniques. *Scientific Reports*, *15*(1), 25890. <https://doi.org/10.1038/s41598-025-11260-y>
- Abuouelezz, W., Ali, N., Aung, Z., Altunaiji, A., Shah, S. B., & Gliddon, D. (2025). Exploring PM2.5 and PM10 ML Forecasting Models: A Comparative Study in the UAE. *Scientific Reports*, *15*(1), 9797. <https://doi.org/10.1038/s41598-025-94013-1>
- Alfasanah, Z., Niam, M. Z. H., Wardiani, S., Ahsan, M., & Lee, M. H. (2025). Monitoring Air Quality Index with EWMA and Individual Charts Using XGBoost and SVR Residuals. *MethodsX*, *14*, 103107. <https://doi.org/10.1016/j.mex.2024.103107>
- Al-Saeedi, K., Zhou, D., Fish, A., Tsakiri, K., & Marsellos, A. (2025). A Methodological Comparison of Forecasting Models Using KZ Decomposition and Walk-Forward Validation. *Mathematics*, *13*(21), 3410. <https://doi.org/10.3390/math13213410>
- Anastasya, N., & Suwandana, E. (2022). Efek Pertumbuhan Ekonomi, Kependudukan, Dan Transportasi Terhadap Kualitas Udara Sumatera Selatan. *Publikasi Penelitian Terapan Dan Kebijakan*, *5*(2). <https://doi.org/10.46774/pptk.v5i2.498>
- Asrul, D. A. I., & Soebroto, A. A. (2024). Optimalisasi Prediksi Kasus COVID-19 di Indonesia: Perbandingan Teknik Validasi 80-20 Split dan Walk-Forward dengan ARIMA. *J-INTECH*, *12*(02), 297–308. <https://doi.org/10.32664/j-intech.v12i02.1373>
- Az-Zahran, M. H., Rasuli, E. H., & Silaban, M. I. S. (2024). Pengembangan Sistem Prediksi Perubahan Iklim Berbasis Kecerdasan Buatan untuk Manajemen Sumber Daya Alam yang Berkelanjutan di Papua. *Jurnal Syntax Admiration*, *5*(9), 3492–3500. <https://doi.org/10.46799/jsa.v5i9.1488>
- Bahadure, N. B., Gonge, S., Parashar, D., Shah, B., Patil, P. D., Renugadevi, M., & Raju, N. (2025). SVR Machine Learning and SARIMA-Based Air Quality Index Classification and Forecasting System. *Discover Applied Sciences*, *7*(9), 994. <https://doi.org/10.1007/s42452-025-07327-0>
- Batool Awan, W., & Ali, Z. (2025). Development of Standardized PM2.5 Concentration Index (SPM2.5I) for Monitoring and Forecasting Air Pollution

- Characteristics. *Theoretical and Applied Climatology*, 156(10), 504. <https://doi.org/10.1007/s00704-025-05798-0>
- Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A Machine Learning Approach to Predict Air Quality in California. *Complexity*, 2020, 1–23. <https://doi.org/10.1155/2020/8049504>
- Chen, Z., Chen, D., Zhao, C., Kwan, M., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., & Xu, B. (2020). Influence of Meteorological Conditions on PM<sub>2.5</sub> Concentrations Across China: A Review of Methodology and Mechanism. *Environment International*, 139, 105558. <https://doi.org/10.1016/j.envint.2020.105558>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Fatima, N., Yousafzai, S. N., Nemri, N., Alsolai, H., Ebad, S. A., Sorour, S., Gu, Y., Syafrudin, M., & Fitriyani, N. L. (2025). Time Series AQI Forecasting Using Kalman-Integrated Bi-GRU and Chi-Square Divergence Optimization. *Scientific Reports*, 15(1), 29157. <https://doi.org/10.1038/s41598-025-12422-8>
- Haya, F., Nisa, K., Ladipasa, R. F., Suriani, A., & Media, A. (2025). Dampak Polusi Udara terhadap Kesehatan Manusia. *WISSEN: Jurnal Ilmu Sosial Dan Humaniora*, 3(2), 180–190. <https://doi.org/10.62383/wissen.v3i2.753>
- Houssein, E. H., Mohamed, M., Younis, E. M. G., & Mohamed, W. M. (2025). A Hybrid Harris Hawks Optimization with Support Vector Regression for Air Quality Forecasting. *Scientific Reports*, 15(1), 2275. <https://doi.org/10.1038/s41598-025-86275-6>
- Joshi, D. C., Negi, P., Devi, S., Lohani, H., Kumar, R., Gupta, M., & Ming, L. C. (2025). Fine Particulate Matter (PM<sub>2.5</sub>, PM<sub>10</sub>): A Silent Catalyst for Chronic Lung Diseases in India; a Comprehensive Review. *Environmental Challenges*, 20, 101215. <https://doi.org/10.1016/j.envc.2025.101215>
- Kalantari, E., Gholami, H., Malakooti, H., Nafarzadegan, A. R., & Moosavi, V. (2024). Machine Learning for Air Quality Index (AQI) Forecasting: Shallow Learning or Deep Learning? *Environmental Science and Pollution Research*, 31(54), 62962–62982. <https://doi.org/10.1007/s11356-024-35404-1>
- Kementerian Lingkungan Hidup dan Kehutanan. (n.d.). *JDIH Kementerian Lingkungan Hidup dan Kehutanan*. Peraturan Lingkungan Hidup Dan

Kehutanan. Retrieved 26 September 2025, from [https://jdih.menlhk.go.id/new2/uploads/files/P\\_14\\_2020\\_ISPU\\_menlhk\\_07302020074834.pdf](https://jdih.menlhk.go.id/new2/uploads/files/P_14_2020_ISPU_menlhk_07302020074834.pdf)

Liu, B., Jin, Y., & Li, C. (2021). Analysis and Prediction of Air Quality in Nanjing from Autumn 2018 to Summer 2019 Using PCR–SVR–ARMA Combined Model. *Scientific Reports*, *11*(1), 348. <https://doi.org/10.1038/s41598-020-79462-0>

Liu, Z., Huang, X., & Wang, X. (2024). PM2.5 Prediction Based on Modified Whale Optimization Algorithm and Support Vector Regression. *Scientific Reports*, *14*(1), 23296. <https://doi.org/10.1038/s41598-024-74122-z>

Maltare, N. N., & Vahora, S. (2023). Air Quality Index Prediction Using Machine Learning for Ahmedabad City. *Digital Chemical Engineering*, *7*, 100093. <https://doi.org/10.1016/j.dche.2023.100093>

Meisenbacher, S., Turowski, M., Phipps, K., Rätz, M., Müller, D., Hagenmeyer, V., & Mikut, R. (2022). Review of Automated Time Series Forecasting Pipelines. *WIREs Data Mining and Knowledge Discovery*, *12*(6), e1475. <https://doi.org/10.1002/widm.1475>

Muhaimin, M. R., Karina, D. M., & Krisna, A. B. (2024). Prediksi Kualitas Udara Malang Menggunakan Metode Gradient Boosting Regression. *Digital Transformation Technology*, *4*(2), 937–942. <https://doi.org/10.47709/digitech.v4i2.5046>

Rahman, Md. M., Nayeem, Md. E. H., Ahmed, Md. S., Tanha, K. A., Sakib, Md. S. A., Uddin, K. M. M., & Babu, H. Md. H. (2024). AirNet: Predictive Machine Learning Model for Air Quality Forecasting Using Web Interface. *Environmental Systems Research*, *13*(1), 44. <https://doi.org/10.1186/s40068-024-00378-z>

Rahman, R. A., White, B., & Ma, C. (2024). The effect of growth, deforestation, forest fires, and volcanoes on Indonesian regional air quality. *Journal of Cleaner Production*, *457*, 142311. <https://doi.org/10.1016/j.jclepro.2024.142311>

Rajesh, M., Babu, R. G., Moorthy, U., & Easwaramoorthy, S. V. (2025). Machine Learning-Driven Framework for Realtime Air Quality Assessment and Predictive Environmental Health Risk Mapping. *Scientific Reports*, *15*(1), 28801. <https://doi.org/10.1038/s41598-025-14214-6>

Riani, M., Atkinson, A. C., & Corbellini, A. (2023). Automatic Robust Box–Cox and Extended Yeo–Johnson Transformations in Regression. *Statistical*

*Methods & Applications*, 32(1), 75–102. <https://doi.org/10.1007/s10260-022-00640-7>

Sarmadi, M., Rahimi, S., Rezaei, M., Sanaei, D., & Dianatinasab, M. (2021). Air Quality Index Variation Before and After the Onset of COVID-19 Pandemic: A Comprehensive Study on 87 Capital, Industrial and Polluted Cities of the World. *Environmental Sciences Europe*, 33(1), 134. <https://doi.org/10.1186/s12302-021-00575-y>

Shafii, N. H. B., Alias, R., Zamani, N. F., & Fauzi, N. F. (2020). Forecasting of Air Pollution Index PM<sub>2.5</sub> Using Support Vector Machine(SVM). *Journal of Computing Research and Innovation*, 5(3), 43–53. <https://doi.org/10.24191/jcrinn.v5i3.149>

Shaykh, 'Abdullah ibn Muhammad ibn 'Abd al-Rahman ibn Ishaq Al, & Ghoffar, M. A. (2006). *Tafsir Ibnu Katsir*. Pustaka Imam Asy-Syafi'i.

Syuhada, G., Akbar, A., Hardiawan, D., Pun, V., Darmawan, A., Heryati, S. H. A., Siregar, A. Y. M., Kusuma, R. R., Driejana, R., Ingole, V., Kass, D., & Mehta, S. (2023). Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia. *International Journal of Environmental Research and Public Health*, 20(4), 2916. <https://doi.org/10.3390/ijerph20042916>

The World Air Quality Index. (n.d.). *BMKG | Badan Meteorologi, Klimatologi Dan Geofisika*. Aqicn.Org. Retrieved 10 September 2025, from <https://aqicn.org/network/bmkg/>

Trong, N. D., Le, T. D., & Nguyen, H. Q. (2026). ATMOS: IoT-Driven Air Quality Monitoring and Forecasting System Using Machine Learning and Fog Computing. In N. T. Nguyen, V. H. Pham, T. D. Tran, T.-P. Hong, Y. Manolopoulos, N. A. Le Khac, & P. T. Tin (Eds), *Computational Intelligence in Engineering Science* (Vol. 2584, pp. 196–208). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-98161-6\\_15](https://doi.org/10.1007/978-3-031-98161-6_15)

Umah, R., & Gusmira, E. (2024). Dampak Pencemaran Udara terhadap Kesehatan Masyarakat di Perkotaan. *Profit: Jurnal Manajemen, Bisnis Dan Akuntansi*, 3(3), 103–112. <https://doi.org/10.58192/profit.v3i3.2246>

United Nations Statistics Division. (2023a). *Metadata – SDG Indicator 3.9.1: Mortality Rate Attributed to Household and Ambient Air Pollution*. United Nations. <https://unstats.un.org/sdgs/metadata/files/Metadata-03-09-01.pdf>

United Nations Statistics Division. (2023b). *Metadata – SDG Indicator 11.6.2: Annual Mean Levels of Fine Particulate Matter (PM<sub>2.5</sub>) in Cities (population*



*Weighted*). United Nations.  
<https://unstats.un.org/sdgs/metadata/files/Metadata-11-06-02.pdf>

U.S. Environmental Protection Agency. (n.d.). *Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)*. Retrieved 26 September 2025, from <https://document.airnow.gov/technical-assistance-document-for-the-reporting-of-daily-air-quality.pdf>

Wong, P.-Y., Zeng, Y.-T., Su, H.-J., Lung, S.-C. C., Chen, Y.-C., Chen, P.-C., Hsiao, T.-C., Adamkiewicz, G., & Wu, C.-D. (2025). Effects of Feature Selection Methods in Estimating SO<sub>2</sub> Concentration Variations Using Machine Learning and Stacking Ensemble Approach. *Environmental Technology & Innovation*, 37, 103996. <https://doi.org/10.1016/j.eti.2024.103996>

Xiang, M., Du, S., Yu, J., & Xu, Y. (2022). Short-Term Load Forecasting Based on Sample Weights Assignment. *Energy Reports*, 8, 783–791. <https://doi.org/10.1016/j.egyr.2022.09.101>

Yunita, R., Gunarto, T., Marselina, M., & Yuliawan, D. (2023). The Influence of GDP Per Capita, Income Inequality, and Population on CO<sub>2</sub> Emission (Environmental Kuznet Curve Analysis in Indonesia). *International Journal of Social Science, Education, Communication and Economics (SINOMICS JOURNAL)*, 2(2), 217–230. <https://doi.org/10.54443/sj.v2i2.130>

Zhang, X., Ma, C., Li, X., Xiong, L., & Nie, S. (2022). Assessing the Impact of Air Pollution on Inbound Tourism Along the Yangtze River Across Space and Time. *International Journal of Environmental Research and Public Health*, 19(17), 10944. <https://doi.org/10.3390/ijerph191710944>

## **LAMPIRAN**

## LAMPIRAN

Lampiran 1 Data Deret Waktu Setelah Rekayasa Fitur (Data *Dummy*)

Janggal	pm25	t	h	p	w	dew	pm25_lagl	t_lagl	h_lagl	p_lagl	w_lagl	dew_lagl	pm25_rmean3	t_rmean3	h_rmean3	p_rmean3	w_rmean3	dew_rmean3	pm25_diff1	pm25_ewma	t_ewma	h_ewma	p_ewma	w_ewma	dew_ewma	dow_sin	dow_cos
24/10/2025	154	25,5	92	1009,1	3	24,1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	154	25,5	92	1009,1	3	24,1	-0,434	-0,901
25/10/2025	159	25,9	95,4	1009,2	2,5	25,1	154	25,5	92	1009,1	3	24,1	NaN	NaN	NaN	NaN	NaN	NaN	5	156,5	25,66	93,36	1009,14	2,8	24,5	-0,975	-0,223
26/10/2025	65	26	100	1009	2	26	159	25,9	95,4	1009,2	2,5	25,1	126	25,8	95,8	1009,13	2,5	24,73	-94	110,75	25,8	96,02	1009,08	2,48	25,1	-0,782	0,623
27/10/2025	69	26,7	95,8	1010	2,1	25,1	65	26	100	1009	2	26	97,67	26,2	97,07	1009,4	2,2	25,4	4	89,88	26,16	95,93	1009,45	2,33	25,1	0	1
28/10/2025	109	26	94	1010	1,5	25	69	26,7	95,8	1010	2,1	25,1	81	26,23	96,6	1009,4	2,2	25,4	40	99,44	26,09	95,16	1009,67	2	25,06	0,782	0,623

Lampiran 2 Hasil Normalisasi Fitur (Data *Dummy*)

Tanggal	pm25'	t'	h'	p'	w'	dew'	pm25_lagl'	t_lagl'	h_lagl'	p_lagl'	w_lagl'	dew_lagl'	pm25_rmean3'	t_rmean3'	h_rmean3'	p_rmean3'	w_rmean3'	dew_rmean3'	pm25_diff1'	pm25_e_wma'	t_ewma'	h_ewma'	p_ewma'	w_ewma'	dew_ewma'	dow_sin'	dow_cos'
24/10/2025	1,06	-1,13	-1,06	-0,61	0,98	-1,1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1,06	-0,62	-0,6	-0,62	0,89	-0,65	-0,434	-0,901
25/10/2025	1,18	-0,69	-0,11	-0,57	0,35	-0,52	1,06	-1,13	-1,06	-0,61	0,98	-1,1	NaN	NaN	NaN	NaN	NaN	NaN	0,11	1,11	-0,49	-0,32	-0,58	0,28	-0,22	-0,975	-0,223
26/10/2025	-1,15	-0,62	1,21	-0,63	-0,26	0,77	1,18	-0,69	-0,11	-0,57	0,35	-0,52	-0,09	-0,54	0,1	-0,59	-0,16	-0,1	-2,33	-0,06	-0,49	0,1	-0,6	-0,16	0,14	-0,782	0,623
27/10/2025	-1,05	0,32	0,35	0,57	-0,1	-0,52	-1,15	-0,62	1,21	-0,63	-0,26	0,77	1,26	0,56	0,4	0,57	0,53	0,37	0,11	-0,54	0,44	0,13	0,59	-0,2	0,14	0	1
28/10/2025	-0,06	-0,62	-0,39	0,63	-0,97	-0,43	-1,05	0,32	0,35	0,57	-0,1	-0,52	-0,49	0,41	0,23	-0,59	-0,31	-0,18	0,99	-0,09	0,35	-0,05	0,75	-0,37	0,07	0,782	0,623