

FINAL PROJECT
EKSPLORASI DAN VISUALISASI DATA
KELAS SD-A2

ANALISIS DATA KUALITAS UDARA DKI JAKARTA PADA TAHUN 2023
DENGAN PENDEKATAN REGRESI BINOMIAL NEGATIF



Kelompok 9

- | | | |
|----|--------------------------|-----------|
| 1. | Evia Ayu Damarani | 164241049 |
| 2. | Jessica Mercy Assa | 164241036 |
| 3. | Mirza Nabila R. H. | 164241013 |
| 4. | Shahnaz Fatharani Azhima | 164241024 |

PROGRAM STUDI TEKNOLOGI SAINS DATA
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN
UNIVERSITAS AIRLANGGA
SURABAYA
2025

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR TABEL.....	4
DAFTAR GAMBAR.....	5
BAB I	
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
BAB II	
TINJAUAN PUSTAKA.....	4
2.1 Tinjauan Pustaka Statistika.....	4
2.1.1 Statistika Deskriptif.....	4
2.1.2 Analisis Regresi Linear.....	4
2.1.3 Analisis Regresi Logistik.....	4
2.1.4 Teknik Pre-Processing Data.....	5
2.1.4.1 Penanganan Missing Values (Data Hilang).....	5
2.1.4.2 Deteksi dan Penanganan Outlier.....	5
2.1.4.3 Regresi Linear Berganda (OLS).....	6
2.1.4.4 Uji Asumsi Klasik.....	6
2.1.4.5 Generalized Linear Model (GLM).....	6
2.1.4.5 Regresi Poisson.....	6
2.1.4.5 Kriteria Pemilihan Model Terbaik.....	7
2.1. Tinjauan Pustaka Non-Statistika.....	7
2.1.1 Pencemaran Udara dan Kesehatan.....	7
2.1.2 Indeks Standar Pencemar Udara (ISPU).....	8
2.1.3 Dinamika Ozon Permukaan (O3), Polutan Lain, dan Faktor Meteorolog....	8
BAB III	
METODOLOGI.....	9
3.1 Sumber Data.....	9
3.2 Metode Pre-Processing.....	10
3.3 Metode Analisis Data.....	10
3.4 Metode Visualisasi Data.....	11
3.4.1 Pie Chart.....	11
3.4.2 Bar Chart.....	11
3.4.3 Box Plot.....	11
3.4.4 Scatter Plot.....	11
BAB IV	
ANALISIS DAN PEMBAHASAN.....	12
4.1 Data Pre-Processing.....	12

4.1.1 Pemeriksaan Struktur dan Tipe Data.....	12
4.1.2 Identifikasi dan Penanganan Missing Values.....	12
4.1.3 Koreksi Tipe Data.....	14
4.1.4 Identifikasi dan Penanganan Nilai Tidak Logis dan Inkonsisten.....	15
4.3 Hasil Analisis.....	16
4.3.1 Statistika Deskriptif.....	16
4.3.2. Regresi Linear.....	17
4.3.2. Regresi GLM Poisson.....	18
4.3.3. Regresi GLM Binomial Negatif.....	19
4.4. Visualisasi Data R-Shiny.....	21
BAB V	
KESIMPULAN DAN SARAN.....	24
5.1 Kesimpulan.....	24
5.2 Saran.....	24
DAFTAR PUSTAKA.....	25
LAMPIRAN.....	29

DAFTAR TABEL

Tabel 3.1 Variabel Penelitian.....	9
Tabel 4.1 Penanganan Missing Values.....	12
Tabel 4.2 Kategori Indeks Standar Pencemar Udara (ISPU).....	13
Tabel 4.3 Perbandingan Nilai R ² Penanganan Outlier.....	16
Tabel 4.4 Ringkasan Pemusatan dan Persebaran Data.....	16
Tabel 4.5 Uji Asumsi Klasik.....	17

DAFTAR GAMBAR

Gambar 4.1 Output Pemeriksaan Struktur dan Tipe Data.....	12
Gambar 4.2 Jumlah Missing Values Sebelum dan Sesudah Penanganan.....	14
Gambar 4.3 Output Pengecekan Tipe Data.....	14
Gambar 4.4 Inkonsistensi Data pada Kolom stasiun.....	15
Gambar 4.5 Boxplot untuk Deteksi Outlier.....	15
Gambar 4.6 Uji Overdispersi.....	19
Gambar 4.7 Uji Regresi Binomial Negatif.....	20
Gambar 4.8 Uji Serentak Regresi Binomial Negatif.....	20
Gambar 4.9 Distribusi Kategori ISPU.....	21
Gambar 4.10 Rata-Rata ISPU per Stasiun.....	22
Gambar 4.11 Tren ISPU Tahun 2023.....	22
Gambar 4.12 Distribusi Konsentrasi Polutan.....	23

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara memberikan dampak yang semakin nyata terhadap ekosistem dan kehidupan manusia. Intergovernmental Panel on Climate Change (IPCC) dalam laporannya menyatakan bahwa aktivitas emisi gas rumah kaca dan polutan udara, merupakan kontributor utama pemanasan global yang mengakibatkan perubahan pola cuaca ekstrem, kenaikan permukaan air laut, dan degradasi kualitas udara (IPCC, 2021). Gas-gas pencemar seperti karbon monoksida (CO), nitrogen dioksida (NO₂), sulfur dioksida (SO₂), dan partikulat matter (PM₁₀) tidak hanya berperan dalam pembentukan gas rumah kaca, tetapi juga menghasilkan ozon troposferik (O₃) yang dapat memperparah perubahan iklim.

Menurut World Health Organization (WHO), paparan jangka panjang terhadap polutan udara dapat meningkatkan risiko penyakit kardiovaskular, gangguan pernapasan, dan mortalitas prematur hingga 7 juta kasus per tahun secara global (WHO, 2022). Oleh karena itu, dunia melalui Sustainable Development Goals (SDGs) ke-13 "*Climate Action*" memberikan perhatian terhadap pentingnya aksi nyata dalam memitigasi perubahan iklim dan dampaknya.

Kawasan perkotaan, sebagai pusat aktivitas ekonomi dan mobilitas, menghadapi tantangan serius terkait pencemaran udara akibat peningkatan pesat jumlah kendaraan bermotor dan ekspansi industri. Di Indonesia, pertumbuhan sektor transportasi mencapai rata-rata 6,8% per tahun, dengan dominasi kendaraan berbahan bakar fosil yang menyumbang sekitar 75% dari total emisi polutan udara di wilayah urban (Kementerian Lingkungan Hidup dan Kehutanan, 2022). Sementara itu, aktivitas industri seperti pembangkit listrik, pabrik manufaktur, dan konstruksi berkontribusi terhadap emisi SO₂ dan PM₁₀ yang dapat bertahan di atmosfer dalam jangka waktu lama.

Data dari IQAir (2023) menempatkan Jakarta sebagai salah satu kota dengan kualitas udara terburuk di Asia Tenggara, dengan konsentrasi $PM_{2.5}$ yang sering melampaui batas aman WHO sebesar $15 \mu g/m^3$. Kontribusi emisi polutan udara di DKI Jakarta didominasi oleh sektor transportasi (44%), industri (31%), dan pembangkit listrik (14%), dengan total emisi mencapai 18.760 ton per hari (Badan Pengelolaan Lingkungan Hidup DKI Jakarta, 2023). Tingginya beban pencemaran ini tidak hanya mengancam kesehatan 10 juta penduduk Jakarta, tetapi juga berkontribusi signifikan terhadap emisi gas rumah kaca nasional.

Paparan terhadap gas-gas pencemar ini menimbulkan beberapa dampak negatif. Bagi manusia, polutan udara dapat menyebabkan iritasi saluran pernapasan, penurunan fungsi paru-paru, hingga peningkatan risiko kanker. Bagi lingkungan, deposisi asam dari SO_2 dan NO_2 merusak ekosistem perairan dan tanah. Bagi iklim, ozon troposferik berperan sebagai gas rumah kaca dengan potensi pemanasan global 1.000-2.000 kali lebih kuat dibandingkan CO_2 dalam jangka pendek (Stevenson et al., 2020). Kondisi ini menunjukkan urgensi pengendalian emisi polutan udara sebagai bagian integral dari strategi mitigasi perubahan iklim dan perlindungan kesehatan publik.

Menyadari urgensi permasalahan ini, pemerintah DKI Jakarta melalui Badan Pengelolaan Lingkungan Hidup Daerah (BPLHD) telah menerapkan sistem pemantauan kualitas udara berbasis Indeks Standar Pencemar Udara (ISPU) sejak tahun 2010. ISPU merupakan indeks yang menggambarkan kondisi kualitas udara di suatu lokasi berdasarkan konsentrasi lima parameter utama: PM_{10} , SO_2 , CO , O_3 , dan NO_2 . Meskipun data pemantauan ISPU telah dikumpulkan secara konsisten selama lebih dari satu dekade, masih terdapat kebutuhan mendesak untuk menganalisis secara mendalam pengaruh antar polutan. Temuan dari penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam memahami dinamika pencemaran udara di Jakarta untuk strategi mitigasi perubahan iklim dan peningkatan kualitas udara di Jakarta dan kota-kota besar lainnya di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan, dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana karakteristik pemusatan dan persebaran Indeks Standar Pencemaran Udara (ISPU) di berbagai stasiun pemantauan DKI Jakarta pada tahun 2023?
2. Bagaimana model yang tepat untuk memodelkan data kualitas udara DKI Jakarta pada tahun 2023?
3. Bagaimana faktor-faktor yang berpengaruh signifikan terhadap parameter Ozon di DKI Jakarta?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui karakteristik pemusatan dan persebaran Indeks Standar Pencemaran Udara (ISPU) di berbagai stasiun pemantauan DKI Jakarta pada tahun 2023.
2. Mengetahui model yang tepat untuk memodelkan data kualitas udara DKI Jakarta pada tahun 2023.
3. Mengetahui faktor-faktor yang berpengaruh signifikan terhadap parameter Ozon di DKI Jakarta.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka Statistika

2.1.1 Statistika Deskriptif

Statistik deskriptif berperan sebagai fondasi awal analisis untuk memahami karakteristik data melalui ukuran pemusatan (Mean, Median) dan penyebaran data (Standard Deviation, Min-Max). Penerapan statistik deskriptif sangat penting untuk memetakan profil distribusi polutan di setiap stasiun pemantauan, mengidentifikasi variabilitas harian, serta mendeteksi indikasi awal adanya anomali sebelum dilakukan pemodelan lebih lanjut. Penggunaan visualisasi deskriptif seperti *box plot* terbukti efektif dalam memberikan ringkasan statistik distribusi data kualitas udara yang sering kali memiliki pola kemiringan (*skewness*) akibat kejadian polusi ekstrem (Rendana & Komariah, 2021).

2.1.2 Analisis Regresi Linear

Analisis Regresi Linear digunakan untuk memodelkan hubungan kuantitatif antar variabel, baik melalui Regresi Linear Sederhana untuk melihat tren satu variabel independen terhadap dependen, maupun Regresi Linear Berganda untuk mengestimasi pengaruh simultan berbagai gas polutan terhadap konsentrasi partikulat dengan memperhatikan asumsi klasik seperti multikolinearitas. Kebaikan model ini dievaluasi menggunakan Koefisien Determinasi (R^2) yang mengukur seberapa besar variabilitas data aktual dapat dijelaskan oleh model. Studi peramalan ISPU di Jakarta menggunakan metode regresi linear tersegmentasi menunjukkan kemampuan model ini dalam memprediksi konsentrasi PM_{2.5} dan PM₁₀ dengan akurasi yang memadai untuk kebutuhan mitigasi polusi. (Sri Aisyah Hidayati, 2025))

2.1.3 Analisis Regresi Logistik

Berbeda dengan prediksi nilai kontinu, Analisis Regresi Logistik dirancang khusus untuk menangani variabel dependen kategorik atau biner,

seperti klasifikasi status udara menjadi "Sehat" atau "Tidak Sehat". Metode ini memanfaatkan fungsi Logit untuk menghitung *Odds Ratio*, yang memberikan interpretasi probabilitas mengenai peluang terjadinya suatu kategori risiko kesehatan berdasarkan kombinasi variabel prediktor. Penerapan *Multinomial Logistic Regression* pada data kualitas udara di Jakarta terbukti mampu menghasilkan klasifikasi kategori ISPU dengan akurasi mencapai 91,67%, terutama ketika variabel PM_{2.5} disertakan sebagai fitur prediktor utama. (Aji dkk., 2025)

2.1.4 Teknik *Pre-Processing* Data

2.1.4.1 Penanganan *Missing Values* (Data Hilang)

Kekosongan data pada sensor lingkungan sering terjadi akibat gangguan teknis, yang dapat dikategorikan sebagai Missing Completely at Random (MCAR) atau Missing at Random (MAR). Untuk data deret waktu yang memiliki pola tren dan musiman yang kuat, metode interpolasi linear atau ARIMA sering digunakan untuk menjaga konsistensi pola temporal. Namun, sebagai alternatif yang lebih kekar (*robust*) terhadap keberadaan pencilan ekstrem, metode Imputasi Median sering lebih disukai; evaluasi komputasi menunjukkan bahwa imputasi Median mampu menghasilkan tingkat kesalahan (*error rate*) yang sebanding dengan metode kompleks seperti Random Forest dalam menangani data PM_{2.5} (Gajić dkk., 2024)

2.1.4.2 Deteksi dan Penanganan *Outlier*

Pencilan dalam data lingkungan didefinisikan sebagai observasi yang menyimpang secara signifikan dari distribusi data mayoritas, yang bisa disebabkan oleh kejadian polusi ekstrem atau kesalahan sensor. Metode *Interquartile Range* (IQR) digunakan untuk mendeteksi pencilan dengan menetapkan batas pagar bawah ($Q1 - 1.5 \times IQR$) dan batas pagar atas ($Q3 + 1.5 \times IQR$) (Maltare dkk., 2025). Pendekatan IQR ini terbukti efektif meningkatkan kinerja model prediksi kualitas udara dengan membersihkan *noise* tanpa menghilangkan informasi variabilitas alami yang penting (Maltare dkk., 2025)

2.1.4.3 Regresi Linear Berganda (OLS)

Analisis Regresi Linear Berganda dengan metode Kuadrat Terkecil (*Ordinary Least Squares/OLS*) digunakan sebagai model dasar (*baseline*) untuk memetakan hubungan linear antara variabel prediktor meteorologis terhadap variabel respon ISPU. Secara matematis, OLS bekerja dengan meminimalkan jumlah kuadrat sisaan (*error*) antara nilai prediksi dan nilai aktual. Studi peramalan ISPU di Jakarta menggunakan regresi linear tersegmentasi menunjukkan bahwa model ini mampu memberikan prediksi yang cukup akurat untuk parameter tertentu, meskipun memiliki keterbatasan dalam menangani data diskrit atau data dengan varian yang tidak konstan (Kamil, 2025).

2.1.4.4 Uji Asumsi Klasik

Agar estimasi OLS bersifat Best Linear Unbiased Estimator (BLUE), beberapa asumsi klasik harus dipenuhi. Uji Normalitas (seperti Kolmogorov-Smirnov) memastikan sisaan menyebar normal. Uji Multikolinearitas menggunakan Variance Inflation Factor (VIF) mendeteksi korelasi antar variabel bebas, di mana nilai $VIF < 10$ mengindikasikan model bebas dari multikolinearitas serius (Cipe dkk., 2025). Uji Homoskedastisitas menggunakan Uji Glejser diperlukan untuk memastikan ragam sisaan konstan; pelanggaran asumsi ini sering terjadi pada data lingkungan. Terakhir, Uji Autokorelasi Durbin-Watson penting untuk data deret waktu guna memastikan tidak ada korelasi antar sisaan pada periode waktu berurutan (Kim, 2025).

2.1.4.5 Generalized Linear Model (GLM)

Generalized Linear Model (GLM) adalah perluasan dari regresi linear yang mengakomodasi variabel respon yang tidak berdistribusi normal, seperti data cacahan (*count data*). GLM menghubungkan rata-rata variabel respon dengan kombinasi linear prediktor melalui fungsi penghubung (*link function*). Pendekatan ini memberikan fleksibilitas lebih besar dibandingkan OLS dalam memodelkan data lingkungan yang sering kali memiliki distribusi miring atau mengikuti keluarga distribusi eksponensial (Saputro, 2021).

2.1.4.5 Regresi Poisson

Regresi Poisson dirancang khusus untuk data respon berupa bilangan cacah (non-negatif). Model ini bekerja di bawah asumsi equidispersion, yaitu

nilai rata-rata ($E(Y)$) sama dengan ragam ($Var(Y)$). Namun, dalam data kualitas udara, sering ditemukan kondisi di mana ragam jauh melebihi rata-rata, yang disebut sebagai fenomena overdispersion. Jika overdispersion diabaikan dalam model Poisson, hal ini dapat menyebabkan estimasi standar error yang bias (terlalu rendah) dan kesimpulan statistik yang tidak valid (Tiara dkk., 2023).

2.1.4.5 Kriteria Pemilihan Model Terbaik

Untuk mengevaluasi dan memilih model terbaik antara OLS, Poisson, dan Binomial Negatif, digunakan kriteria *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC). Kedua metrik ini menilai keseimbangan antara kebaikan suai (*goodness of fit*) dan kompleksitas model, dengan memberikan penalti pada penambahan parameter yang tidak perlu. Dalam pemodelan deret waktu polutan udara, model dengan nilai AIC dan BIC terkecil dianggap sebagai model yang paling optimal karena mampu memberikan prediksi akurat dengan struktur yang paling efisien (prinsip parsimoni).

2.1. Tinjauan Pustaka Non-Statistika

Bagian ini membahas teori dasar mengenai variabel lingkungan dan kerangka regulasi yang melandasi penelitian dataset ISPU di DKI Jakarta.

2.1.1 Pencemaran Udara dan Kesehatan

Pencemaran udara didefinisikan sebagai kontaminasi lingkungan atmosfer oleh agen kimia, fisik, atau biologis yang mengubah karakteristik alami atmosfer. Di wilayah perkotaan padat seperti DKI Jakarta, urgensi pemantauan kualitas udara sangat krusial karena polusi udara telah menjadi faktor risiko utama penyakit tidak menular kronis, seperti gangguan kardiovaskular dan pernapasan. Studi terbaru mencatat bahwa Jakarta sering menempati peringkat atas kota dengan kualitas udara terburuk secara global, yang mendorong perlunya intervensi kebijakan berbasis data prediksi yang akurat untuk memitigasi dampak kesehatan masyarakat (Kamil, 2024).

2.1.2 Indeks Standar Pencemar Udara (ISPU)

Pemerintah Indonesia mengatur komunikasi risiko kualitas udara melalui Indeks Standar Pencemar Udara (ISPU) yang landasan hukumnya mengacu pada Peraturan Menteri Lingkungan Hidup dan Kehutanan (PermenLHK) Nomor 14 Tahun 2020. Regulasi ini mengonversi konsentrasi polutan terukur menjadi angka tanpa satuan yang dikelompokkan ke dalam kategori: Baik (0–50), Sedang (51–100), Tidak Sehat (101–200), Sangat Tidak Sehat (201–300), dan Berbahaya (>300) (Aji dkk., 2025). Penentuan status mutu udara harian didasarkan pada "parameter pencemar kritis", yaitu parameter polutan dengan nilai ISPU tertinggi pada saat pengukuran dilakukan (PermenLHK, 2020 dalam Aji dkk., 2025).

2.1.3 Dinamika Ozon Permukaan (O_3), Polutan Lain, dan Faktor Meteorolog

Ozon permukaan (O_3) merupakan polutan sekunder yang terbentuk di troposfer melalui reaksi fotokimia kompleks antara Nitrogen Oksida (NO_x) dan Senyawa Organik Volatil (VOC) dengan bantuan intensitas radiasi matahari, menjadikannya sangat sensitif terhadap rasio prekursor serta fluktuatif (Lee et al., 2024; Zhang et al., 2022). Namun, dinamika kualitas udara tidak berdiri sendiri; interaksi dengan parameter lain seperti partikulat halus ($PM_{2.5}$, PM_{10}) serta gas buang (CO , NO_2 , SO_2) memegang peranan vital, di mana analisis *feature importance* di Jakarta sering menempatkan partikulat sebagai determinan utama klasifikasi ISPU (Irfayana dkk., 2025). Selain komposisi kimiawi, aspek temporal juga berpengaruh melalui konsep *lag effect*, di mana akumulasi polutan dari hari sebelumnya dapat terus berdampak pada kondisi kesehatan dan mutu udara saat observasi akibat sifat persistensi atmosferik (Kardinah Lestari dkk., 2019).

BAB III METODOLOGI

3.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh melalui *website* resmi Satu Data Jakarta. Data ini diterbitkan oleh Dinas Lingkungan Hidup dengan judul Data Indeks Standar Pencemar Udara (ISPU) di Provinsi DKI Jakarta 2023. Data tersebut mencakup variabel-variabel relevan untuk menganalisis kualitas udara di 5 stasiun DKI Jakarta dengan rentang tahun 2023. Ozon (O₃) dipilih sebagai variabel dependen yang menggambarkan indeks kualitas udara dengan parameter ozon berdasarkan indikator penilaian udara bersih. Variabel independen yang digunakan adalah PM 10, PM 2.5, sulfur dioksida (SO₂), karbon monoksida (CO), dan nitrogen dioksida (NO₂) sebagai indikator indeks kualitas udara dari parameter partikulat udara lainnya, stasiun sebagai indikator nama Stasiun Pemantau Kualitas Udara (SPKU) tempat pelaksanaan pengukuran ISPU, serta kategori sebagai indikator kategori hasil pengukuran ISPU tertinggi pada SPKU dan tanggal tertentu.

Tabel 3.1 Variabel Penelitian

No	tanggal	stasiun	pm _{sepeuluh}	pm _{duaakomalingga}	sulfur _{dioksida}	karbon _{monoksida}	ozon	nitrogen _{dioksida}	max	parameter _{pencemar_kritis}	kategori
1	2023-02-25	DKI5 Kebon Jeruk Jakarta Barat	35	-	13	12	31	18	35	PM10	BAIK
2	2023-02-26	DKI5 Kebon Jeruk Jakarta Barat	23	-	14	9	32	11	32	O3	BAIK
...
1824	2023-11-29	DKI5 Kebon Jeruk	56	88	32	18	56	18	88	PM25	SEDANG
1825	2023-11-30	DKI5 Kebon Jeruk	30	57	32	15	61	17	61	O3	SEDANG

3.2 Metode *Pre-Processing*

Data *Pre-Processing* adalah langkah awal yang dilakukan setelah pengambilan data untuk membersihkan dan mempersiapkan *raw* data agar siap digunakan dalam analisis dan juga pembuatan model. Pada penelitian ini, 2 langkah yang dilakukan pada data *pre-processing* adalah mendeteksi dan menangani *missing value* serta mendeteksi *outlier*. Untuk mendeteksi *missing values*, digunakan *function* `is.na()`. *Function* ini digunakan untuk mengidentifikasi elemen-elemen yang memiliki *missing value* yang ditangani dengan 2 metode yaitu *interpolasi* dan imputasi median yang disempurnakan dengan penanganan inkonsistensi data.

Kemudian, kami menggunakan *boxplot* yang untuk mengidentifikasi adanya nilai yang berada di luar batas normal pada data. *Outlier* ditandai sebagai titik yang berada di luar batas ($Q3 + 1.5 \times IQR$) atau batas bawah ($Q1 - 1.5 \times IQR$). *Interquartile Range* (IQR) digunakan untuk menentukan rentang data yang dianggap normal dengan menghitung selisih antara kuartil pertama (Q1) dan kuartil ketiga (Q3).

3.3 Metode Analisis Data

Metode Analisis Data merupakan serangkaian langkah sistematis yang bertujuan untuk mengeksplorasi dan menganalisis informasi yang terdapat dalam suatu dataset. Proses ini dirancang untuk memperoleh wawasan, mengidentifikasi pola, dan menggali informasi berharga dari data yang telah terkumpul. Dalam penelitian ini, regresi menjadi metode analisis data yang digunakan pada penelitian ini. Regresi menjadi metode analisis data yang digunakan pada penelitian ini.

Regresi merupakan sebuah alat yang digunakan untuk memperkirakan nilai dari suatu variabel yang tidak diketahui berdasarkan satu atau beberapa variabel yang sudah diketahui (Pradita, 2022). Model regresi *poisson* merupakan model regresi nonlinear yang digunakan untuk menganalisis data diskrit (*count*). Dalam regresi *poisson* terdapat asumsi *equidispersi*, namun sering terjadi pelanggaran seperti kasus *overdispersi*. Analisis regresi binomial negatif merupakan alternatif analisis yang digunakan ketika asumsi *equidispersion* pada

analisis regresi *poisson* tidak terpenuhi atau disebut dengan *overdispersion* (Lestari, 2019).

3.4 Metode Visualisasi Data

3.4.1 *Pie Chart*

Pie chart biasanya dapat digunakan untuk analisis data agar dapat melihat distribusi proporsional dari suatu kategori. Dalam hal ini, metode *pie chart* dapat menganalisis terkait proporsi harian berdasarkan kategori Indeks Standar Pencemar Udara (ISPU), seperti Baik, Sedang, Tidak Sehat, atau Berbahaya selama periode pengamatan.

3.4.2 *Bar Chart*

Bar chart biasanya dapat digunakan untuk analisis data agar dapat melakukan perbandingan antarkelompok data secara visual dan terstruktur. Dalam hal ini, metode *bar chart* dapat menganalisis terkait perbedaan rata-rata konsentrasi ISPU antarstasiun pengamatan atau untuk melihat tren ISPU harian sepanjang waktu.

3.4.3 *Box Plot*

Box Plot merupakan ringkasan distribusi sampel yang disajikan secara grafis yang bisa menggambarkan bentuk distribusi data (*skewness*), ukuran tendensi sentral dan ukuran penyebaran (keragaman) data pengamatan (*Mengenal Box-Plot (Box and Whisker Plots) – Accounting, 2020*). Dengan menggunakan *box plot*, dapat menyimpulkan bahwa data yang digunakan memiliki outlier atau tidak.

3.4.4 *Scatter Plot*

Scatter plot biasanya dapat digunakan untuk analisis data agar dapat melihat hasil data yang lebih jelas dan akurat. Dalam hal ini, metode *scatter plot* dapat menganalisis terkait korelasi antara nilai aktual vs prediksi konsentrasi O_3 menggunakan model binomial negatif

BAB IV

ANALISIS DAN PEMBAHASAN

4.1 Data Pre-Processing

4.1.1 Pemeriksaan Struktur dan Tipe Data

```

Jumlah baris: 1825
Jumlah kolom: 12

periode_data      tanggal
"character"      "character"
stasiun           pm_sepuluh
"character"      "character"
pm_duakomalima    sulfur_dioksida
"character"      "character"
karbon_monoksida  ozon
"character"      "character"
nitrogen_dioksida max
"character"      "character"
parameter_pencemar_kritis kategori
"character"      "character"

```

Gambar 4.1 Output Pemeriksaan Struktur dan Tipe Data

Data terdiri dari 1825 baris dan 12 kolom yang semuanya masih dalam tipe data *character*. Kolom tersebut adalah *periode_data*, *tanggal*, *stasiun*, *pm_sepuluh*, *pm_duakomalima*, *sulfur_dioksida*, *karbon_monoksida*, *ozon*, *nitrogen_dioksida*, *max*, *parameter_pencemar_kritis*, dan *kategori*.

4.1.2 Identifikasi dan Penanganan *Missing Values*

Sebelum mengidentifikasi *missing values*, terdapat beberapa string seperti “-” dan “---” yang diubah terlebih dahulu menjadi NA agar dikenali sebagai *missing values*. Berikut ringkasan kolom yang memiliki *missing values* dan penanganannya:

Tabel 4.1 Penanganan *Missing Values*

Kolom	Jumlah	Penanganan
pm_sepuluh	222	1) Imputasi dengan interpolasi karena data <i>time series</i> . 2) Imputasi dengan median dari tanggal dan parameter (pm_sepuluh, pm_duakomalima, sulfur_dioksida,
pm_duakomalima	295	
sulfur_dioksida	39	
karbon_monoksida	44	

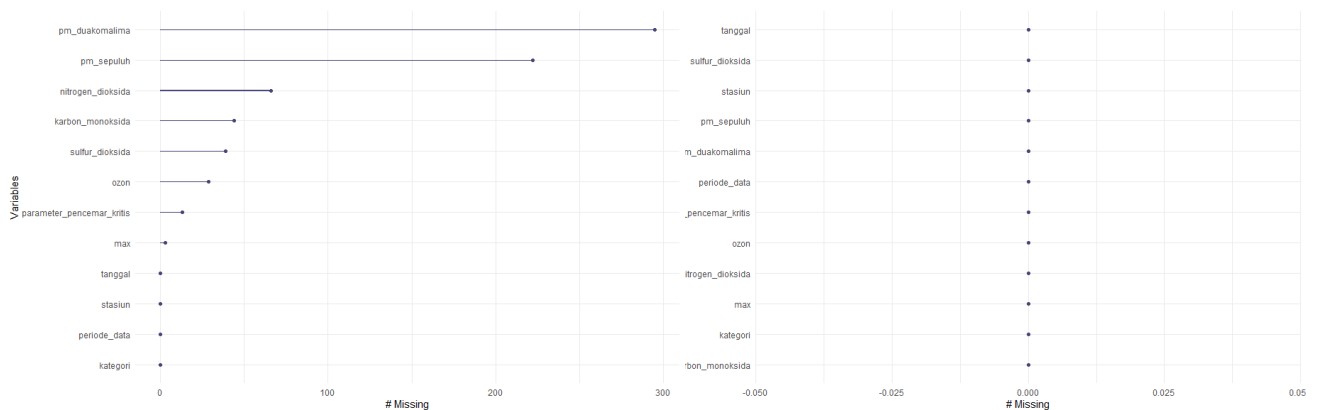
ozon	29	karbon_monoksida, nitrogen_dioksida) ozon, yang sama.
nitrogen_dioksida	66	
max	3	Memperbarui data dengan nilai <i>maximum</i> parameter di tanggal dan stasiun yang sama.
parameter_pencemar_kritis	13	Memperbarui data dengan parameter yang memiliki nilai <i>maximum</i> parameter di tanggal dan stasiun yang sama.

Selain itu, terdapat kelompok “TIDAK ADA DATA” pada kolom kategori sehingga nilainya diperbarui berdasarkan tabel Konversi Nilai Konsentrasi yang bersumber dari Kementerian Lingkungan Hidup dan Kehutanan.

Tabel 4.2 Kategori Indeks Standar Pencemar Udara (ISPU)

Rentang	Kategori
1-50	Baik
51-100	Sedang
101-200	Tidak Sehat
201-300	Sangat Tidak Sehat
301+	Berbahaya

Kemudian, dilakukan visualisasi perbandingan sebelum dan sesudah penanganan *missing values*.



Gambar 4.2 Jumlah *Missing Values* Sebelum dan Sesudah Penanganan

4.1.3 Koreksi Tipe Data

```
## List of 12
## $ periode_data      : chr [1:12] "202212" "202301" "202302" "202303" "202304" "202305" "202306" "202307" "202308" "202309" "202310" "202311"
## $ tanggal          : Date[1:365], format: "2022-12-01" "2022-12-02" ...
## $ stasiun           : Factor w/ 5 levels "DKI1 Bunderan HI",...: 1 2 3 4 5
## $ pm_sepuluh        : num [1:88] 54 59 53 64 52 55 60 48 56 49 58 44 65 61 45 74 66 51 70 40 50 41 28 3
5 42 36 43 25 24 32 37 18 29 26 47 20 27 46| __truncated__
## $ pm_duakomalima    : num [1:131] 73 74 81 93 78 67 92 94 80 76 100 75 70 71 95 117 79 86 107 84 68 72
88 62 63 85 58 69 91 42 61 54 53 45 43 38 57| __truncated__
## $ sulfur_dioksida   : num [1:53] 36 46 42 52 18 39 51 43 20 38 50 47 35 45 17 49 16 34 33 44 48 21 30 3
1 41 15 32 23 19 57 22 53 54 62 59 24 58 25| __truncated__
## $ karbon_monoksida  : num [1:47] 12 15 9 7 17 11 14 18 10 8 16 13 19 24 21 6 27 20 5 3 1 29 30 26 28 22
23 45 49 38 4 2 55 54 42 41 25 34 37 36 33| __truncated__
## $ ozon              : num [1:72] 22 50 24 27 14 42 20 21 49 29 26 32 45 18 28 48 13 41 17 37 47 25 34 3
1 12 10 23 15 16 8 30 11 19 38 33 46 39 7 6| __truncated__
## $ nitrogen_dioksida : num [1:49] 13 31 12 19 5 30 18 11 17 34 14 33 10 44 16 40 35 9 2 41 8 7 6 32 22 4
23 15 3 1 45 48 47 24 27 26 21 20 25 0 29 | __truncated__
## $ max               : num [1:127] 73 74 81 93 78 67 92 94 80 76 100 75 70 71 95 117 79 86 107 84 68 72
88 62 63 85 58 69 91 42 61 54 53 45 43 41 57| __truncated__
## $ parameter_pencemar_kritis: chr [1:4] "pm_duakomalima" "sulfur_dioksida" "pm_sepuluh" "ozon"
## $ kategori          : chr [1:4] "SEDANG" "TIDAK SEHAT" "BAIK" "SANGAT TIDAK SEHAT"
```

Gambar 4.3 Output Pengecekan Tipe Data

Secara keseluruhan, struktur dataset sudah hampir sesuai untuk analisis statistik lanjutan karena beberapa variabel, seperti tanggal, parameter pencemar udarasudah ditangani sebelum menangani *missing values*. Kemudian, tipe data kategori dan parameter pencemaran kritis diubah menjadi *factor*.

4.1.4 Identifikasi dan Penanganan Nilai Tidak Logis dan Inkonsisten

Pada data ini, ada beberapa nilai inkonsisten pada kolom stasiun, yaitu adanya penggunaan spasi ganda dan penamaan ganda untuk satu jenis stasiun yang sama sehingga dilakukan perbaikan dan perubahan tipe data ke *factor*.

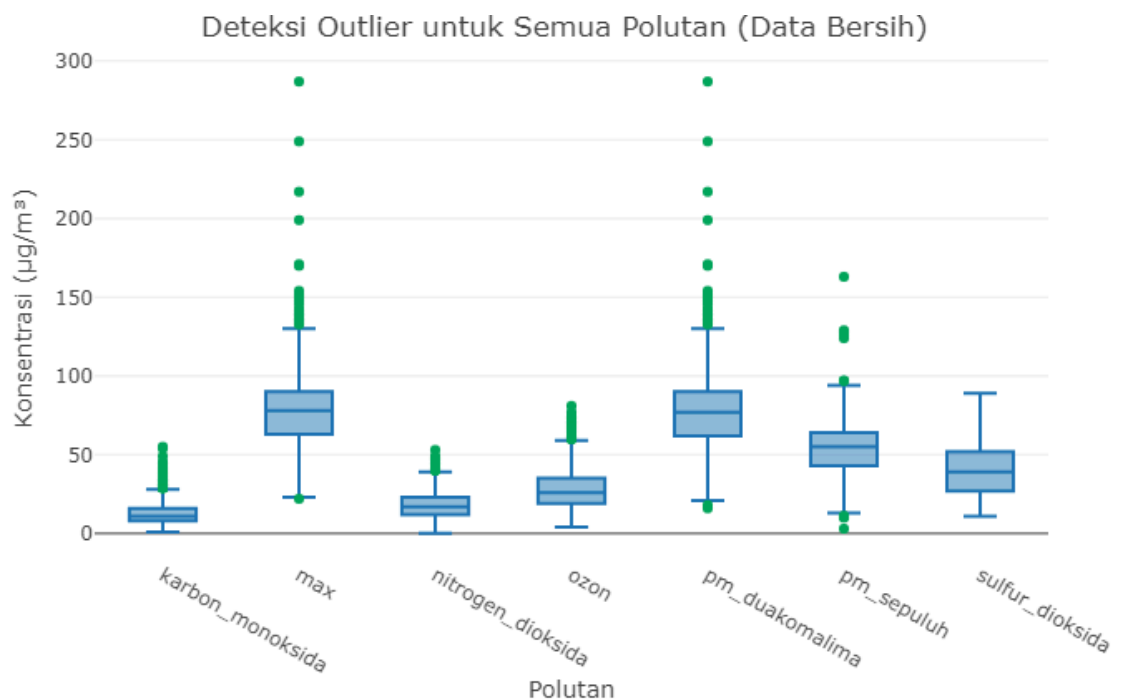
```
$ stasiun      : chr [1:10] "DKI5 Kebon Jeruk Jakarta Barat" "DKI1 Bunderan HI" "DKI2  
Kelapa Gading" "DKI3 Jagakarsa" "DKI4 Lubang Buaya" "DKI1 Bunderan HI" "DKI2 Kelapa Gading" "DKI3  
Jagakarsa" "DKI5 Kebon Jeruk" "DKI4 Lubang Buaya"
```

Gambar 4.4 Inkonsistensi Data pada Kolom stasiun

4.1.5 Pembuatan Variabel Baru

Berdasarkan kolom tanggal, kami mengekstrak tanggal, bulan, dan tahun sebagai variabel-variabel baru untuk dimasukkan ketika melakukan pembuatan model.

4.1.6 Deteksi dan Penanganan *Outlier* dan *Influential Observations*



Gambar 4.5 Boxplot untuk Deteksi *Outlier*

Untuk mendeteksi *outlier*, pertama kami menggunakan visualisasi boxplot. Kemudian, kami membuat model sebelum dilakukan penanganan *outlier*. Setelah itu, kami membuat model tanpa influential observation dengan menghapus observasi yang melebihi pengukuran Leverage (rule of thumb 3 *h*), Studentized Result (rule of thumb 2), dan Cook's distance (rule of thumb 1). Kemudian, kami membuat model dengan menghapus outlier berdasarkan rumus ± 1.5 IQR. Berikut perbandingannya:

Tabel 4.3 Perbandingan Nilai R^2 Penanganan *Outlier*

Model	Multiple R^2	Adjusted R^2
Model tanpa penanganan outlier	0.6801	0.6774
Model setelah menghapus influential observation	0.7667	0.7647
Model setelah menghapus outlier IQR	0.6850	0.6824

Dari tabel tersebut, kami memutuskan bahwa penanganan yang paling tepat adalah dengan menghapus *influential observation* saja untuk mendapatkan model dengan R^2 tertinggi.

4.3 Hasil Analisis

4.3.1 Statistika Deskriptif

Tabel 4.4 Ringkasan Pemusatan dan Persebaran Data

Variabel	Min.	Q1	Median	Mean	Q3	Max.	Std.
tanggal	2022-12-01	2023-03-08	2023-06-04	2023-06-04	2023-09-01	2023-11-30	15.08
pm_sepuluh	10	44	55	53.32	64	129	21.82
pm_duakoma lima	16	62	78	77.03	91	171	13.18
sulfur_dioksida	11	27	39	38.59	52	79	6.15

karbon_monoksida	1	8	11	12.23	16	43	12.07
ozon	5	20	26	27.61	34	65	8.57
nitrogen_dioksida	0	12	17	17.4	23	53	20.68
max	22	63	78	77.84	91	171	0.25
tahun	2022	2023	2023	2023	2023	2023	0.25
bulan	1	3	6	6.394	9	12	3.38
hari	1	8	16	15.67	23	31	8.75

Tabel tersebut memperlihatkan pemusatan data dari data yang sudah dilakukan *pre-processing* data. Pada data tersebut, dapat kita ketahui bahwa perhitungan dimulai dari 1 Desember 2022 hingga 30 November 2023. Selain itu, beberapa parameter, seperti sulfur dioksida, karbon monoksida, ozon, dan nitrogen dioksida yang memiliki mean dengan rentang 0-50 yang berarti baik. Sedangkan parameter pm 10 dan pm 2.5 memiliki mean dengan rentang 51-100 yang berada dalam kategori sedang. Adapun persebaran dari data ini tidak melebihi mean dan median sehingga dapat dikatakan tersebar dengan baik. Dari sini pula, kita dapat mengetahui bahwa data memiliki skala data diskrit dan non-negatif.

4.3.2. Regresi Linear

Sebelum melakukan uji regresi linear berganda, kami melakukan uji asumsi klasik dengan ringkasan hasil sebagai berikut:

Tabel 4.5 Uji Asumsi Klasik

Asumsi	Uji	Statistik Uji	<i>p-value</i>	Keputusan	Interpretasi
Linearitas	Reset test	41.067	< 2.2e-16	Tolak H0	Variabel dependen dan variabel independen tidak linear

Normalitas	Kolmogorov-Smirnov	0.017317	0.709	Gagal Tolak H0	<i>Error</i> berdistribusi normal
Multikolinearitas	VIF	< 10		Gagal Tolak H0	Variabel tidak berkorelasi tinggi
Heteroskedastisitas	Glejser	170.6988	4.947107e-29	Tolak H0	Variansi <i>error</i> tidak konstan (heteroskedastisitas)
Autokorelasi	Durbin-Watson	1.5429	< 2.2e-16	Tolak H0	<i>Error</i> dependen (ada autokorelasi)

Karena belum memenuhi uji linearitas, heteroskedastisitas, dan autokorelasi, maka analisis tidak bisa menggunakan model *Ordinary Least Square* (OLS) untuk melakukan estimasi parameter.

4.3.2. Regresi GLM Poisson

Salah satu cara untuk mengatasi pelanggaran uji asumsi heteroskedastisitas dan autokorelasi adalah dengan menggunakan model Generalized Linear Model (GLM) lainnya. Setelah mengamati bentuk data yang memiliki tipe data diskrit dan non-negatif, maka kami melakukan analisis menggunakan model regresi poisson.

Untuk melakukan regresi poisson, terdapat beberapa prasyarat yang harus dipenuhi, yaitu:

- 1) Tidak ada multikolinearitas pada variabel independen (sudah terpenuhi)
- 2) Nilai variansi variabel respons sama dengan mean

Uji Overdispersi

H_0 : $\phi = 1$ (Data tidak memiliki overdispersi)

H_0 : $\phi > 1$ (Data memiliki overdispersi)

```

Overdispersion test

data: model_poisson
z = 7.6172, p-value = 1.296e-14
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
1.286053

```

Gambar 4.6 Uji Overdispersi

Berdasarkan uji overdispersi tersebut, dapat diketahui bahwa p-value ($2.667e^{-14}$) $> \alpha$ sehingga diambil keputusan tolak H_0 yang berarti terdapat overdispersi pada data. Hal ini berarti asumsi bahwa nilai variansi variabel respons sama dengan mean tidak terpenuhi.

4.3.3. Regresi GLM Binomial Negatif

Untuk mengatasi overdispersi, digunakan regresi binomial negatif dengan output sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \beta_k = 0$ (Variabel prediktor secara bersama-sama tidak berpengaruh signifikan terhadap variabel respons)

$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p$ (Minimal ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respons)


```

Call:
glm.nb(formula = ozon ~ pm_sepuluh + pm_duakomalima + sulfur_dioksida +
  karbon_monoksida + nitrogen_dioksida + stasiun + kategori +
  tahun + bulan + hari, data = data_model_nb, init.theta = 117.4236483,
  link = log)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.025e+03  6.052e+01 -16.934 < 2e-16 ***
pm_sepuluh      3.432e-04  7.282e-04   0.471  0.63743
pm_duakomalima  5.612e-03  5.960e-04   9.416 < 2e-16 ***
sulfur_dioksida -7.871e-04  5.471e-04  -1.439  0.15024
karbon_monoksida -3.980e-03  1.327e-03  -3.000  0.00270 **
nitrogen_dioksida 4.148e-03  1.052e-03   3.942  8.07e-05 ***
stasiunDKI2 Kelapa Gading 1.986e-01  1.777e-02  11.176 < 2e-16 ***
stasiunDKI3 Jagakarsa -7.478e-02  2.442e-02  -3.063  0.00219 **
stasiunDKI4 Lubang Buaya -3.304e-01  2.704e-02 -12.220 < 2e-16 ***
stasiunDKI5 Kebon Jeruk  5.543e-01  2.486e-02  22.293 < 2e-16 ***
kategoriSEDANG  2.555e-02  2.774e-02   0.921  0.35696
kategoriTIDAK SEHAT -3.863e-02  4.057e-02  -0.952  0.34102
tahun          5.078e-01  2.992e-02  16.972 < 2e-16 ***
bulan          5.120e-02  2.103e-03  24.347 < 2e-16 ***
hari           2.182e-03  6.064e-04   3.599  0.00032 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(117.4236) family taken to be 1)

Null deviance: 7131.8 on 1669 degrees of freedom
Residual deviance: 1775.0 on 1655 degrees of freedom
AIC: 10604

Number of Fisher Scoring iterations: 1

      Theta: 117.4
Std. Err.: 22.0

2 x log-likelihood: -10571.54

```

Gambar 4.7 Uji Regresi Binomial Negatif

Berdasarkan model regresi binomial negatif menggunakan variabel prediktor tersebut didapatkan persamaan sebagai berikut:

$$\hat{Y} = \exp(-1025 + 0.0003432(\text{PM } 10) + 0.005612(\text{PM } 2.5) - 0.0007871(\text{SO}_2) - 0.00398(\text{CO}) + 0.004148(\text{NO}_2) + 0.1986(\text{stasiun DKI2}) - 0.07478(\text{stasiun DKI3}) - 0.3304(\text{stasiun DKI4}) + 0.5543(\text{stasiun DKI5}) + 0.02555(\text{kategori SEDANG}) - 0.03863(\text{kategori TIDAK SEHAT}) + 0.5078(\text{tahun}) + 0.05120(\text{bulan}) + 0.002182(\text{hari}))$$

```

Response: ozon

1
2 pm_sepuluh + pm_duakomalima + sulfur_dioksida + karbon_monoksida + nitrogen_dioksida + stasiun + kategori + tahun + bulan + hari
   theta Resid. df    2 x log-lik.    Test    df LR stat. Pr(Chi)
1   6.256018   1669    -12913.93      G     14 2342.395      0
2 117.423648   1655    -10571.54  1 vs 2

```

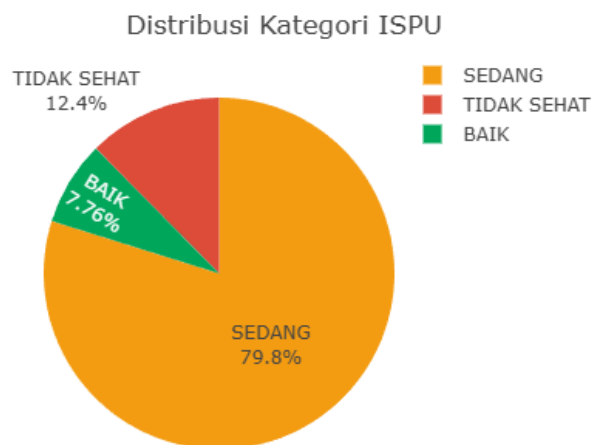
Gambar 4.8 Uji Serentak Regresi Binomial Negatif

Berdasarkan uji serentak menggunakan Maximum Likelihood Ratio Test, didapatkan $p\text{-value}$ (hampir mendekati 0) $< \alpha$ dan LR stat. $(2342.395) > X^2_{df=14, \alpha=0.05} (23.68)$ sehingga diambil keputusan tolak H_0 yang berarti minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel respons.

Berdasarkan uji parsial, variabel prediktor dengan $p\text{-value} < \alpha$ dapat diputuskan tolak H_0 atau memiliki pengaruh signifikan. Variabel yang berpengaruh signifikan terhadap Ozon adalah:

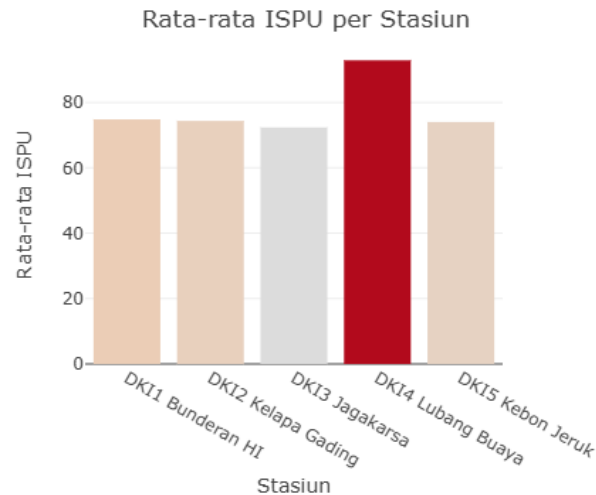
- PM 2.5
- Karbon monoksida (CO)
- Nitrogen dioksida (NO₂)
- Stasiun DKI 2 Kelapa Gading
- Stasiun DKI 3 Jagakarsa
- Stasiun DKI 4 Lubang Buaya
- Stasiun DKI 5 Kebon Jeruk
- Tahun
- Bulan
- Hari

4.4. Visualisasi Data R-Shiny



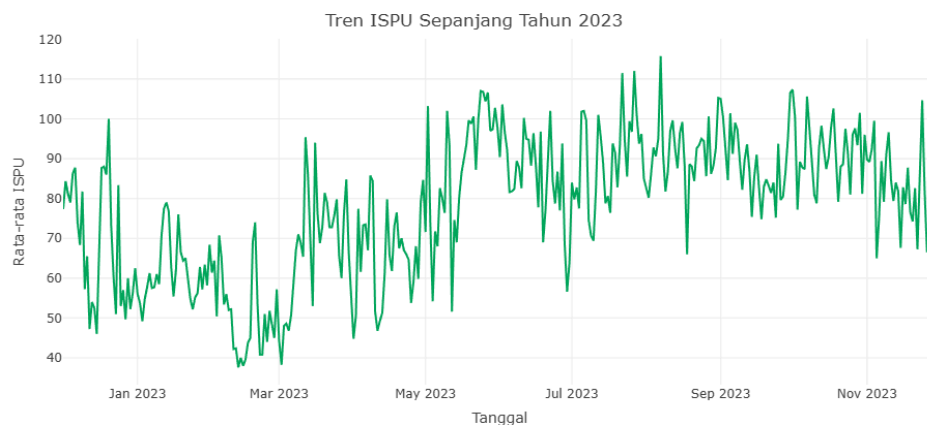
Gambar 4.9 Distribusi Kategori ISPU

Rata-rata kualitas udara berada pada kategori sedang (79,8%), menunjukkan udara cukup aman namun tetap berisiko. Kategori tidak sehat tercatat 12,4%, dan kategori baik hanya 7,76%, sehingga kondisi udara yang benar-benar bersih masih jarang terjadi.



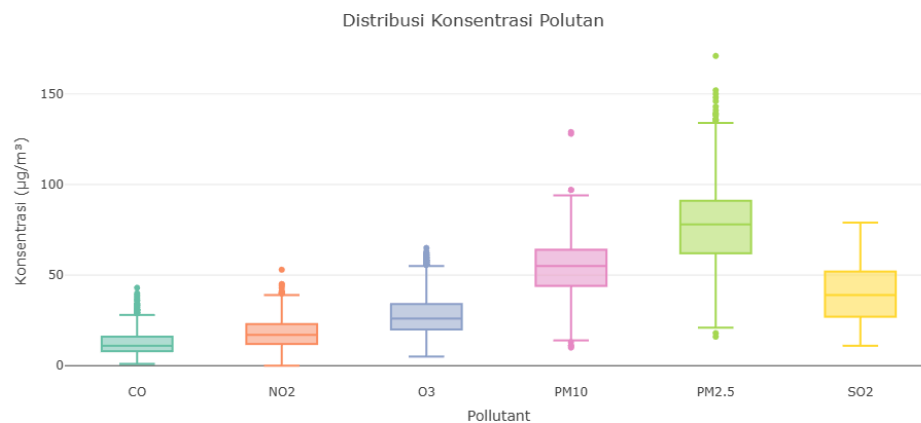
Gambar 4.10 Rata-Rata ISPU per Stasiun

Stasiun DK14 Lubang Buaya memiliki rata-rata ISPU paling tinggi dibandingkan stasiun lainnya, menunjukkan tingkat pencemaran udara yang lebih serius di area tersebut. Sementara stasiun DK11 Bunderan HI, DK12 Kelapa Gading, dan DK15 Kebon Jeruk berada pada tingkat yang relatif serupa, namun lebih tinggi dibanding DK13 Jagakarsa.



Gambar 4.11 Tren ISPU Tahun 2023

Grafik menunjukkan bahwa kualitas udara sepanjang tahun 2023 mengalami kenaikan yang cukup tinggi. Pada awal tahun, nilai ISPU relatif lebih rendah dengan beberapa penurunan hingga kisaran 40–60. Memasuki pertengahan tahun (Mei–Agustus), terjadi peningkatan ISPU yang lebih sering mencapai 90–110, menandakan penurunan kualitas udara pada periode tersebut. Pada akhir tahun, nilai ISPU tetap naik namun pada level menengah hingga tinggi. Secara keseluruhan, tren menunjukkan peningkatan variabilitas dan kecenderungan ISPU yang lebih tinggi pada pertengahan tahun 2023.



Gambar 4.12 Distribusi Konsentrasi Polutan

Grafik menunjukkan bahwa setiap polutan memiliki sebaran konsentrasi yang berbeda. CO dan NO2 memiliki konsentrasi relatif rendah dan stabil, dengan rentang yang sempit. O3 menunjukkan variasi yang sedikit lebih besar dibanding dua polutan sebelumnya. PM10 dan terutama PM2.5 memiliki konsentrasi paling tinggi dan variasi paling lebar, menandakan fluktuasi besar dan potensi risiko kualitas udara yang lebih signifikan. SO2 berada pada tingkat menengah dengan variasi sedang.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Secara umum, kualitas udara di DKI Jakarta didominasi oleh kategori "SEDANG" (79,8%), dengan PM2.5 sebagai parameter kritis utama yang mencatatkan konsentrasi rata-rata 77,03, jauh lebih tinggi dibandingkan PM10 (53,32), dengan stasiun DKI3 (Jagakarsa) mencatatkan observasi tertinggi dan DKI5 (Kebon Jeruk) terendah.
2. Model Regresi Binomial Negatif terbukti menjadi model terbaik untuk memodelkan data kualitas udara (Ozon) di DKI Jakarta 2023, dengan nilai AIC dan BIC terendah dibandingkan model Regresi Poisson, serta kemampuannya menangani overdispersion (nilai dispersi $1.286 > 1$).
3. Analisis model menunjukkan bahwa PM2.5, Karbon Monoksida (CO), Nitrogen Dioksida (NO₂), serta faktor waktu dan lokasi stasiun pemantauan (DKI2 Kelapa Gading dan DKI5 Kebon Jeruk) mempengaruhi konsentrasi Ozon, sementara Sulfur Dioksida (SO₂) dan PM10 tidak menunjukkan pengaruh signifikan.

5.2 Saran

1. Mengingat PM2.5 adalah parameter kritis dominan, kebijakan pengendalian emisi kendaraan bermotor dan industri harus diprioritaskan, terutama di wilayah dengan konsentrasi tinggi seperti yang terpantau di stasiun-stasiun kritis.
2. Disarankan bagi peneliti selanjutnya atau pemerintah untuk menggunakan model yang mengakomodasi *overdispersion* seperti Binomial Negatif dalam prediksi kualitas udara, serta mempertimbangkan variabel meteorologi (suhu, kelembapan) yang belum tercakup dalam studi ini untuk meningkatkan akurasi prediksi.

DAFTAR PUSTAKA

- Aji, A. D., Aswandi, H., & Syarifuddin, H. (2025). *Analysis of the Relationship Between Aerosol Optical Depth (AOD) and Particulate Concentration (PM_{2.5}) for Determining the Standard Index of Air Pollution in Jambi Province*. 2025.
<https://doi.org/10.54543/kesans.v4i10.398>
- Badan Pengelolaan Lingkungan Hidup DKI Jakarta. (2023). Status Lingkungan Hidup Daerah Provinsi DKI Jakarta Tahun 2022. Jakarta: BPLHD DKI Jakarta.
https://lingkunganhidup.jakarta.go.id/uploads/images/galery/iklh-2022_20250702125548.pdf
- Dongre, P. K., Patel, V., Bhoi, U., & Maltare, N. N. (2025). An outlier detection framework for Air Quality Index prediction using linear and ensemble models. *Decision Analytics Journal*, 14. <https://doi.org/10.1016/J.DAJOUR.2025.100546>
- Ghaida, A., Firdaus, F. M., Qatrunnada, K. M., Peters, D., Cardenas, B., & Lestari, P. (2024). Spatial patterns of PM_{2.5} air pollution in Jakarta: insights from mobile monitoring. *E3S Web of Conferences*, 485.
<https://doi.org/10.1051/E3SCONF/202448506002>
- Huboyo, H. S., Samadikun, B. P., Manullang, O. R., & Khasanah, N. (2025). Assimilative Capacity of Air Pollutants Using Emission Inventory and Dispersion Model. *Jurnal Rekayasa Kimia & Lingkungan*, 20(1), 184–195.
<https://doi.org/10.23955/RKL.V20I1.43810>
- IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
<https://www.ipcc.ch/report/ar6/wg1>
- Irjayana, R. C., Fadlil, A., & Umar, R. (2025). Air Quality Index Classification: Feature Selection for Improved Accuracy with Multinomial Logistic Regression. *Jurnal Teknik Informatika (Jutif)*, 6(5), 3265–3279.
<https://doi.org/10.52436/1.JUTIF.2025.6.5.5155>

- Kamil, M. (2024). *FORECASTING THE AIR QUALITY INDEX (AQI) IN JAKARTA, INDONESIA BY USING A SEGMENTED MULTIPLE LINEAR REGRESSION MODEL*. <https://doi.org/10.13140/RG.2.2.34971.89122>
- Kementerian Lingkungan Hidup dan Kehutanan. (2020). Indeks Standar Pencemar Udara (ISPU) sebagai Informasi Mutu Udara Ambien di Indonesia. Jakarta: KLHK. <https://ditppu.menlhk.go.id/portal/read/indeks-standar-pencemar-udara-ispu-sebagai-i-informasi-mutu-udara-ambien-di-indonesia>
- Kardinah Lestari, N., Wahyu Hikmawan, D., Safitrah, I., & Widodo, E. (2019). *ANALISIS FAKTOR LINGKUNGAN PENYEBAB KASUS DIARE DI JAWA BARAT TAHUN 2016 MENGGUNAKAN REGRESI BINOMIAL NEGATIF*.
- Kim, H. (2022). A finite sample correction for the panel Durbin–Watson test. *Applied Economics*, 54(28), 3197–3205. <https://doi.org/10.1080/00036846.2020.1869172>
- Lestari, N. K. (2019). Analisis Faktor Lingkungan Penyebab Kasus Diare di Jawa Barat Tahun 2016 menggunakan Regresi Binomial Negatif. *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) IV 2019*, 1(1), 1–10. <http://hdl.handle.net/11617/11111>
- Mengenal Box-Plot (Box and Whisker Plots) – Accounting*. (2020, December 19). BINUS Accounting. Retrieved November 23, 2025, from <https://accounting.binus.ac.id/2020/12/19/mengenal-box-plot-box-and-whisker-plots/>
- Pradita, N. (2022). Penerapan Data Mining sebagai Cara untuk Memprediksi Prestasi Siswa Berdasarkan Status Ekonomi dan Kedisiplinan Menggunakan Metode Regresi Linier Berganda. *Jurnal Ilmiah Mahasiswa Manajemen, Bisnis dan Akuntansi*, 4(5), 683–696. <https://doi.org/10.32639/jimmba.v4i5.181>
- Qonitan, F. D., Kuncoro, N. W., Rahman, A., Rahmi, S. A., & Damayanti, S. (2024b). Characterizing PM2.5 Pollution in Urban Jakarta: Insights from Morphology, Elemental Composition, and Concentration Analysis. *Springer Proceedings in Physics*, 305, 87–99. https://doi.org/10.1007/978-981-97-0740-9_9

- Rendana, M., & Komariah, L. N. (2021). The relationship between air pollutants and COVID-19 cases and large-scale social restriction's impact on the air quality in Jakarta, Indonesia. *Jurnal Pengelolaan Sumberdaya Alam Dan Lingkungan*, 11(1), 93–100. <https://doi.org/10.29244/JPSL.11.1.93-100>
- Saputro, D. R. S., Susanti, A., & Pratiwi, N. B. I. (2021). The handling of overdispersion on Poisson regression model with the generalized Poisson regression model. *AIP Conference Proceedings*, 2326(1). <https://doi.org/10.1063/5.0040330/1000550>
- Sri Aisyah Hidayati. (2025). Linear Regression Analysis between PM2.5 Exposure Levels and Low Birth Weight (LBW) Incidence at Regional General Hospitals in DKI Jakarta. *Miracle Get Journal*, 2(3), 1–9. <https://doi.org/10.69855/MGJ.V2I3.209>
- Stevenson, D. S., et al. (2020). Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmospheric Chemistry and Physics*, 20(5), 3113–3139. <https://doi.org/10.5194/acp-13-3063-2013>
- Tiara, Y., Aidi, M. N., Erfiani, & Rachmawati, R. (2023). OVERDISPERSION HANDLING IN POISSON REGRESSION MODEL BY APPLYING NEGATIVE BINOMIAL REGRESSION. *Barekeng*, 17(1), 417–426. <https://doi.org/10.30598/BAREKENGVOL17ISS1PP0417-0426>
- World Health Organization. (2022). WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. Geneva: WHO. <https://www.who.int/publications/i/item/9789240034228>

LAMPIRAN

Lampiran 1: *Link* RShiny: shinyapps.io

Lampiran 2: Dataset & Code: bit.ly/UASEVDKe19