Jurnal Sains dan Teknologi Lingkungan Volume 17, Number 2, June 2025 Hal. **167-187**
p-ISSN:2085-1227 dan e-ISSN:2502-6119

KLASIFIKASI DAN PEMETAAN SPASIAL KUALITAS UDARA BERBASIS ISPU MENGGUNAKAN SUPPORT VECTOR MACHINE DENGAN DATA MULTISUMBER

Destiana Salsabila[1)], Nurul Fadila Utami [1)*]
[1)] Politeknik Statistika STIS, Jakarta Timur, Indonesia
*Korespondensi: 222212810@stis.ac.id

Abstrak

Indonesia, khususnya DKI Jakarta, menghadapi tantangan berat pada permasalahan polusi udara dengan konsentrasi PM2.5 yang jauh di atas ambang batas yang ditetapkan oleh WHO, menempatkan Jakarta di peringkat atas kota-kota dengan polusi udara tertinggi. Penelitian ini bertujuan untuk mendapatkan gambaran umum tentang kondisi udara DKI Jakarta, klasifikasi kualitas udara menggunakan algoritma SVM, dan menghasilkan peta kualitas udara beresolusi tinggi dengan mengintegrasikan data dari stasiun pemantauan udara (SPKU) dan citra satelit. Metode yang digunakan adalah imputasi dan resampling SMOTE-Tomek Link untuk mengatasi masalah data hilang dan ketidakseimbangan kelas. Hasil dari penelitian ini menunjukkan bahwa pencemaran udara di DKI Jakarta berfluktuasi sepanjang tahun 2022-2024. Data citra satelit memiliki proporsi data hilang lebih tinggi, model klasifikasi berbasis data ini tetap menunjukkan performa optimal. Sebaliknya, data dari SPKU mengalami peningkatan signifikan pada performa setelah dilakukan *resampling*. Peta spasial yang dihasilkan dari data satelit memberikan resolusi yang lebih tinggi dan menunjukkan nilai ISPU untuk wilayah yang tidak terjangkau SPKU.

Kata kunci: Citra Satelit, ISPU, Klasifikasi, SMOTE-Tomek Link, SVM

Abstract

Indonesia, particularly DKI Jakarta, faces severe challenges regarding air pollution, with PM2.5 concentrations far exceeding the thresholds set by the WHO, placing Jakarta among the top cities with the highest levels of air pollution. This study aims to provide a comprehensive overview of air quality conditions in DKI Jakarta, classify air quality using SVM algorithm, and generate high-resolution air quality maps by integrating data from air quality monitoring stations (SPKU) and satellite imagery. The methodology involves imputation and SMOTE-Tomek Link resampling techniques to address missing data and class imbalance problems. The results indicate that air pollution levels in DKI Jakarta fluctuate throughout the 2022–2024 period. Satellite imagery data contains more missing values, however, classification models based on these data still demonstrate optimal performance. In contrast, data from SPKU shows significant improvement in model performance after resampling. Spatial maps derived from satellite data offer finer resolution and can estimate air quality in areas beyond SPKU coverage.

Keywords: *AQI, Classification, Satellite imagery, SMOTE-Tomek Link, SVM*

1. PENDAHULUAN

Polusi udara kini semakin menjadi ancaman global serius yang dapat mempengaruhi kesehatan manusia bahkan kelestarian lingkungan. Polusi udara adalah masuknya berbagai zat berbahaya ke dalam atmosfer bumi, baik yang disebabkan oleh tindakan manusia maupun kejadian alam (Kurniawan, Sulistiyanti, & Murdika, 2023). Polusi udara menyebabkan jutaan kematian dini setiap tahunnya. Sekitar 89% dari kematian dini tersebut terjadi di negara-negara berpenghasilan rendah dan menengah, dan jumlah terbesar terjadi di wilayah Asia Tenggara dan Pasifik Barat (WHO, 2024). Berdasarkan data dari IQAir, pada tahun 2024 rata-rata

Submitted: 27 July 2024
Accepted: 20 June 2025

konsentrasi PM2.5 di Indonesia mencapai 7,1 kali lipat dari nilai pedoman tahunan PM2.5 yang ditetapkan oleh WHO (IQAir, 2024). Hal itu diperkuat dengan laporan dari World Air Quality 2024, dalam daftar 20 besar negara dengan udara paling tercemar, Indonesia menempati peringkat ke-15 dengan rata-rata konsentrasi PM2.5 sebesar 35,5 $\mu g/m^3$. Meski angka ini turun sekitar 4 persen dibandingkan tahun sebelumnya, Indonesia tetap menjadi negara dengan polusi udara tertinggi di Asia Tenggara (IQAir, 2024). Jakarta sebagai Ibu Kota negara Indonesia menduduki peringkat kedua kota dengan rata-rata konsentrasi PM 2.5 per tahun tertinggi pada tahun 2024 setelah Kota Bekasi (IQAir, 2024). PM atau (*Particulate Matter*) adalah indikator proksi yang umum digunakan untuk polusi udara. Komponen utama PM adalah sulfat, nitrat, amonia, natrium klorida, karbon hitam, debu mineral, dan air. Sebagai upaya mitigasi dari masalah tersebut, diperlukan adanya analisis lebih lanjut mengenai polusi udara salah satunya adalah pengembangan model klasifikasi tingkat polusi udara yang nantinya dapat digunakan sebagai dasar ilmiah dalam membuat kebijakan lingkungan yang efektif dan tepat sasaran (Han, Zhang, Liu, & Xiong, 2018).

Di Indonesia tingkat kualitas udara dapat diklasifikasikan menjadi 5 kategori yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya. Kategori ini didasarkan pada hasil nilai ISPU (Indeks Standar Pencemaran Udara). ISPU adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi mutu udara di suatu wilayah tertentu. ISPU didasarkan pada dampak yang ditimbulkan terhadap kesehatan manusia dan makhluk hidup lainnya (Pontoh et al., 2023). Pada tahun 2020, Kementerian Lingkungan Hidup dan Kehutanan (KLHK) mengeluarkan Peraturan Menteri No. 14 Tahun 2020 tentang Indeks Standar Pencemar Udara (ISPU), yaitu tentang Perhitungan, Pelaporan, dan Informasi Standar Pencemar Udara (KLHK, 2025). Berdasarkan peraturan tersebut terdapat 7 polutan udara yang termasuk di dalam penghitungan nilai ISPU diantaranya adalah Hidrokarbon (HC), Karbon Monoksida (CO), Sulfur Dioksida (SO_2), Nitrogen Dioksida (NO_2), Ozon (O_3), dan Partikulat (PM10 dan PM2.5). Namun, pada penelitian ini mengecualikan data HC karena keterbatasan sumber data. Data yang digunakan dalam penelitian ini bersumber dari stasiun pemantauan udara (SPKU) yang merupakan bagian dari infrastruktur kota DKI Jakarta. Meskipun demikian, stasiun-stasiun ini memiliki keterbatasan dalam hal cakupan spasial serta memerlukan biaya tinggi untuk pemasangan dan pemeliharaannya, sehingga kurang mampu merepresentasikan sebaran polutan secara menyeluruh di wilayah yang lebih luas (Han et al., 2018). Oleh karena itu, pengindraan jarak jauh turut dimanfaatkan sebagai solusi alternatif yang mampu memberikan informasi kualitas udara dalam skala yang lebih luas.

Data polusi udara baik dari citra satelit maupun stasiun pemantauan udara memiliki potensi dalam hal data hilang dan ketidakseimbangan data dimana kedua masalah tersebut secara signifikan dapat mempengaruhi kinerja model klasifikasi sehingga sangat penting untuk ditangani. Sebagai upaya untuk mengatasi masalah tersebut maka dilakukan imputasi menggunakan imputasi dengan median. Imputasi median memiliki keuntungan dalam menangani pencilan dalam data yang diamati dan distribusi data yang tidak seimbang. Selain data yang hilang, masalah lain yang perlu diatasi adalah ketidakseimbangan kelas. Ketidakseimbangan kelas dapat ditangani menggunakan teknik resampling. Pada penelitian ini menggunakan SMOTE-Tomek Link yang merupakan salah satu teknik *resampling* yang menggabungkan SMOTE sebagai metode *oversampling* dan Tomek Links sebagai metode *undersampling* (Chandra, Suprihatin, & Resti, 2023). Keuntungan dari metode SMOTE-Tomek Links adalah dapat meningkatkan ketidakseimbangan data lebih efektif daripada SMOTE dan dapat meningkatkan akurasi kelas minoritas.

Sebagai pendekatan alternatif yang menjanjikan, *machine learning* telah membuka peluang besar dalam mengatasi permasalahan polusi udara. Berbeda dengan metode tradisional yang memerlukan pemahaman menyeluruh terhadap mekanisme internal polusi udara, teknik *machine learning* mampu menangkap pola dan hubungan yang mendasari langsung dari data historis, sehingga menghasilkan prediksi yang akurat dengan kebutuhan komputasi yang relatif lebih rendah (Chandra et al., 2023). Salah satu model *machine learning* yang populer adalah Support Vector Machine atau SVM. SVM adalah seperangkat metode pembelajaran yang diawasi terkait yang digunakan untuk klasifikasi dan regresi (Hovi, Hadiana, & Umbara, 2022). Pada penelitian ini menggunakan SVM karena dinilai mampu mengklasifikasikan kualitas udara dengan performansi akurasi yang sangat tinggi yaitu bisa mencapai 95% sampai 99% bergantung pada node sensor yang digunakan (Handayani, Salamah & Wibowo, 2020). Penelitian lain juga menyebutkan bahwa klasifikasi kualitas udara di Jakarta dengan menggunakan algoritma SVM memiliki kinerja lebih baik dibandingkan dengan algoritma KNN (Jayadi, Pratama, & Wibowo, 2023). Untuk menentukan *hyperparameter* yang optimal dapat menggunakan bantuan Grid Search. Grid Search adalah metode pencarian *hyperparameter* sistematis yang mengevaluasi setiap kombinasi parameter pada ruang pencarian yang telah ditentukan untuk mengoptimalkan kinerja model (Nugraha dan Sasongko, 2022).

Pemetaan spasial kualitas udara beresolusi tinggi menjadi sangat penting mengingat keterbatasan cakupan spasial dari stasiun pemantauan udara yang umumnya hanya mencakup

area terbatas di wilayah perkotaan seperti DKI Jakarta. Ketergantungan pada data titik dari stasiun pemantauan tidak cukup untuk merepresentasikan variasi konsentrasi polutan yang bersifat dinamis dan heterogen. Oleh karena itu, dengan menggabungkan data dari stasiun pemantauan dan citra satelit, pemetaan kualitas udara dapat dilakukan secara lebih menyeluruh dan detail. Pendekatan ini telah dibuktikan dalam penelitian oleh Rowley dan Karakuş (2022) menggunakan pendekatan multimodal AI untuk menghasilkan peta kualitas udara resolusi tinggi di Eropa berdasarkan kombinasi data Sentinel-5P dan data monitoring darat.

Berdasarkan penjelasan latar belakang di atas, terdapat tiga tujuan utama dalam penelitian ini. Pertama memperoleh gambaran umum terkait kondisi polusi udara di DKI Jakarta. Kedua, Melakukan pembersihan data dan menggunakan algoritma Support Vector Machine untuk mengklasifikasikan kualitas udara DKI Jakarta berbasis ISPU menggunakan sumber data tradisional dan citra satelit. Ketiga, menghasilkan pemetaan spasial kualitas udara beresolusi tinggi dengan menggabungkan distribusi data titik stasiun pemantauan udara

2. METODE PENELITIAN

Dalam penelitian ini unit analisis penelitian, yaitu Provinsi DKI Jakarata yang terdiri dari 5 kota yaitu Jakarta Pusat, Jakarta Barat, Jakarta Utara, Jakarta Selatan, dan Jakarta Timur. Data yang digunakan berupa data polutan udara periode 2022 hingga Maret 2025. Data tahun 2022–2024 dimanfaatkan untuk analisis deskriptif guna menggambarkan sebaran spasial masing-masing variabel polutan. Selanjutnya, data periode Januari 2024–Maret 2025 digunakan sebagai data input dalam pengembangan model *supervised learning*. Pengolahan analisis data dalam penelitian ini menggunakan *software* QGIS dan Python.

2.1. Data dan Sumber Data

Tabel 1. Variabel Penelitian

Nama Variabel	Produk	Resolusi Spasial	Satuan
Aerosol Optical Depth (AOD)	MODIS/061/MCD19A2_GRANULES	1000 meter	Adimensional
Sulfur Dioxide (SO_2)	COPERNICUS/S5P/OFFL/L3_SO2	1113.2 meter	mol/m^2
Nitrogen Dioxide (NO_2)	COPERNICUS/S5P/OFFL/L3_NO2	1113.2 meter	mol/m^2
Ozone (O_3)	COPERNICUS/S5P/OFFL/L3_O3	1113.2 meter	mol/m^2
Carbon Monoxide (CO)	COPERNICUS/S5P/OFFL/L3_CO	1113.2 meter	mol/m^2

Sumber: Google Earth Engine

Data dikumpulkan dari dua sumber yang berbeda yaitu data kualitas udara tradisional yang diambil dari stasiun pemantauan kualitas udara (SPKU) yang tersebar di lima titik. Dataset tersebut diperoleh dari situs web Satu Data Jakarta yang dapat diunduh melalui alamat berikut (https://data.jakarta.go.id/). Sementara itu, data citra satelit dapat diperoleh dengan mengambil data dari masing-masing satelit yaitu satelit Sentinel-5P dan MODIS. Sentinel-5P dipilih karena merupakan satelit yang memiliki produk utama berupa data atmosfer polutan udara (Mohammad et al., 2025). Kedua sumber data di atas bersifat komplementer, SPKU berperan sebagai *ground truth* dengan akurasi lokal, sedangkan Sentinel-5P dan MODIS menyediakan cakupan spasial yang lebih luas untuk analisis sebaran polutan pada skala yang lebih luas. Selanjutnya dilakukan proses filter pada data citra dengan melakukan *date selection*, *feature selection*, *region based selection*, *mean reducing*, *clipping area*, *scaling*, dan *flattening* dengan AOI (*Area of Interest*) DKI Jakarta. Untuk menentukan konsentrasi PM10 dan PM2.5 dapat menggunakan data AOD karena AOD memiliki korelasi kuat dengan PM10 (Rui-Jun et al 2019). Selain itu, semakin banyak studi yang menggunakan data AOD untuk menentukan konsentrasi PM2.5 menggunakan beragam model dengan hasil yang positif salah satunya dengan model regresi (Nguyen et al., 2015). Berikut adalah formula yang digunakan untuk menghitung PM2.5 dan PM10 menggunakan data AOD:

$$PM10 = 40{,}948 + 0{,}042AOD \quad (1)$$

$$PM2.5 = 54{,}216 + 0{,}035AOD \quad (2)$$

2.2. Praproses Data

Setelah data dikumpulkan, tahapan selanjutnya yang dilakukan adalah praproses data. Proses awal yang dilakukan adalah imputasi data hilang menggunakan metode median imputation. Metode ini menghitung nilai median bukan berdasarkan data aktual, melainkan menghitung median untuk tiap-tiap kelas berdasarkan kategori ISPU. Data yang tidak seimbang dapat menyebabkan bias pada klasifikasi yang menyebabkan kecenderungan ke kelas mayoritas sehingga diperlukan *preprocessing* untuk mengatasi hal tersebut. Salah satu metode yang dapat digunakan adalah Synthetic Minority Oversampling Technique (SMOTE). SMOTE menghasilkan sampel baru dengan melakukan interpolasi acak antara fitur sampel untuk tiap target dan tetangga terdekatnya (Chandra et al., 2023). Teknik ini meningkatkan jumlah kelas minoritas dan meningkatkan kemampuan generalisasi model. Penghitungan sampel sintesis dilakukan dengan rumus:

$$x_{s_i} = x_{c_i} + r(x_{c_i} - x_{k_i}) \quad (3)$$

Proses dihentikan ketika data untuk setiap kelas telah seimbang. Selanjutnya dilakukan Tomek Links dengan memilih sepasang sampel dengan jarak *Euclidean* minimum dari k tetangga terdekat, di mana setiap sampel berasal dari kelas yang berbeda (x_g, x_h). Proses ini berhenti ketika keseimbangan kelas telah tercapai. Pasangan sampel (x_g, x_h). adalah Tomek Link, jika tidak ada sampel x_k yang memenuhi kondisi berikut:

$$d(x_g, x_k) < d(x_g, x_h) \;\; or \;\; d(x_h, x_k) < \; d(x_g, x_h) \tag{4}$$

Kemudian untuk data citra satelit dilakukan konversi satuan polutan dari kandungan pencemar C_{col} dengan satuan (mol/m^2) ke $(\mu g/m^3)$ agar dapat dibandingkan dengan standar ambang batas nasional (Savenets, 2021). Zat pencemar udara menyebar di lapisan bawah troposfer dengan ketinggian yang berbeda-beda (H). Selain itu, dibutuhkan juga informasi masa molar setiap polutan (g/mol) serta Konstanta (A) sama dengan 1000000. Berikut adalah formula yang digunakan untuk melakukan konversi satuan:

$$C = \frac{C_{col}}{H} \; M.A \tag{5}$$

2.3. Pelabelan Data Menggunakan ISPU

Sebelum melakukan klasifikasi menggunakan metode *supervised learning*, data terlebih dahulu harus dilakukan pelabelan. Seperti yang sudah dijelaskan sebelumnya, pada penelitian ini akan menggunakan aturan nilai ISPU untuk mengkategorikan tingkat klasifikasi polusi udara. Pada penelitian ini proses perhitungan ISPU mengikuti cara yang telah ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020 Tentang ISPU. Berikut adalah persamaan dan rincian nilai untuk konversi nilai konsentrasi parameter ISPU (KLHK, 2025).

$$I \; = \frac{I_a - I_b}{X_a - X_b} (X_x - \; X_b) + \; I_b \tag{6}$$

I : Nilai ISPU

I_a : Batas atas nilai ISPU

I_b : Batas bawah nilai ISPU

X_a : Batas atas dari konsentrasi ambien

X_b : Batas bawah dari konsentrasi ambien

X_x : Hasil pengukuran dari konsentrasi ambien nyata

Setelah didapatkan nilai ISPU nilai tersebut kemudian dilakukan pengelompokan ke dalam 5 kategori kualitas udara yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya. Berikut dilampirkan kategori angka rentang hasil perhitungan nilai ISPU:

Tabel 2. Kategori angka rentang ISPU

Kategori	Status Warna	Angka Rentang
Baik	Hijau	1-50
Sedang	Biru	51-100
Tidak Sehat	Kuning	101-200
Sangat Tidak Sehat	Merah	201-300
Berbahaya	Hitam	≥ 301

Sumber:Peraturan Menteri Lingkungan Hidup dan Kehutanan NO. P.14/MENLHK/SETJEN/KUM.1/7/2020

2.4. Model Klasifikasi Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dalam pembelajaran mesin (*machine learning*) yang termasuk ke dalam *supervised learing*. SVM bekerja dengan cara mencari *hyperplane* optimal yang memisahkan data ke dalam dua kelas yang berbeda dengan margin maksimum. Dalam konteks klasifikasi, SVM bertujuan untuk memaksimalkan jarak antara dua kelas data terdekat dari masing-masing kelas terhadap *hyperplane* tersebut. Data yang berada paling dekat dengan *hyperplane* ini disebut sebagai *support vectors*, yang berperan penting dalam menentukan posisi dan orientasi *hyperplane* (Cortes dan Vapnik, 1995). Salah satu keunggulan utama dari SVM adalah kemampuannya untuk bekerja pada data berdimensi tinggi dan menangani masalah klasifikasi yang tidak linear. Untuk menangani data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel seperti linear, polynomial, radial basis function/RBF yang memetakan data ke dalam ruang berdimensi lebih tinggi agar data dapat dipisahkan secara linier dalam ruang tersebut.

Fungsi kernel memungkinkan SVM untuk menangani kompleksitas distribusi data tanpa perlu melakukan transformasi manual terhadap fitur. Selain itu, terdapat parameter C dan gamma yang digunakan sebagai input ke SVM dan mempengaruhi proses optimasi yang membagi *hyperplane* dengan cara yang optimal berdasarkan data pelatihan (Ben-Hur dan Weston 2010). Parameter C berfungsi sebagai pengatur besar kecilnya penalti terhadap kesalahan klasifikasi pada data. Semakin tinggi nilai C, semakin besar penalti yang dikenakan terhadap kesalahan, sehingga model cenderung menghasilkan *hyperplane* yang lebih ketat dengan lebih sedikit toleransi terhadap data yang salah klasifikasi. Sebaliknya, nilai C yang lebih rendah memberikan toleransi lebih besar terhadap kesalahan, yang memungkinkan *hyperplane* memisahkan kelas dengan margin yang lebih lebar. Sementara itu, parameter gamma menentukan seberapa jauh pengaruh satu titik data terhadap pemisahan kelas. Gamma yang rendah menghasilkan model dengan batas keputusan yang mendekati linear, sedangkan gamma yang tinggi memungkinkan model membentuk batas yang lebih kompleks dan melengkung.

Namun, jika nilai gamma terlalu tinggi, model bisa terlalu menyesuaikan diri dengan data pelatihan (*overfitting*), sehingga performanya menurun pada data baru.

Untuk memperoleh model klasifikasi yang paling optimal, diperlukan penentuan nilai parameter terbaik berdasarkan dataset yang digunakan. Oleh karena itu, dalam penelitian ini digunakan metode grid search untuk mengeksplorasi kombinasi parameter C, gamma, dan kernel. Grid search bekerja dengan menjelajahi seluruh ruang parameter yang telah ditentukan sebelumnya, kemudian mengevaluasi performa model untuk setiap kombinasi melalui proses validasi silang (*cross-validation*). Metode ini memungkinkan pemilihan parameter yang memberikan kinerja terbaik secara empiris, sehingga meningkatkan generalisasi model terhadap data baru.

2.5. Evaluasi Model

Untuk menentukan model terbaik yang digunakan penelitian ini menggunakan evaluasi dari nilai *accuracy*, *precision*, *recall*, dan *F1-Score*. *Accuracy* dalam klasifikasi adalah persentase kepadatan baris data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil. *Precision* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar. *F1-Score* adalah rata-rata harmonik dari *precision* dan *recall*. *Confusion matrix* merupakan tabel dengan empat istilah yang merupakan representasi dari hasil proses klasifikasi pada *confusion matrix* yaitu *true positif*, *true negatif*, *false positif*, dan *false negatif*. Berikut dilampirkan formula dari *accuracy*, *precision*, *recall*, dan *F-1 Score* berdasarkan *confusion matrix*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+ FP} \quad (8)$$

$$Recall = \frac{TP}{TP+ FN} \quad (9)$$

$$F1 - Score = \frac{2\ x\ Presisi\ x\ Recall}{Presisi+Recall} \quad (10)$$

TP: *True Positive* (Model memprediksi data di kelas positif dan data sebenarnya di kelas positif)

FP: *False Positive* (Model memprediksi data di kelas positif, namun data sebenarnya di kelas negatif)

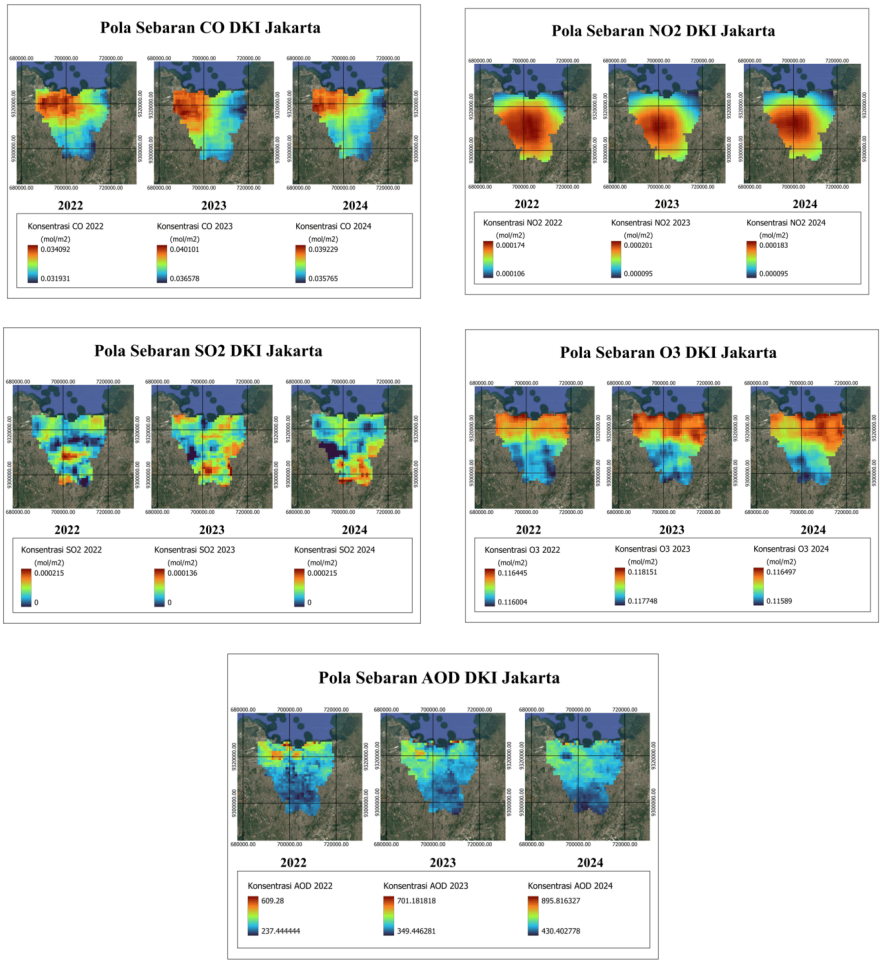
TN: *True Negatif* (Model memprediksi data di kelas negatif dan sebenarnya di kelas negatif)

FN: *False Negative* (Model memprediksi data di kelas negatif, namun data sebenarnya di kelas positif)

3. HASIL DAN PEMBAHASAN

3.1. Gambaran Umum Kondisi Polusi Udara Di DKI Jakarta

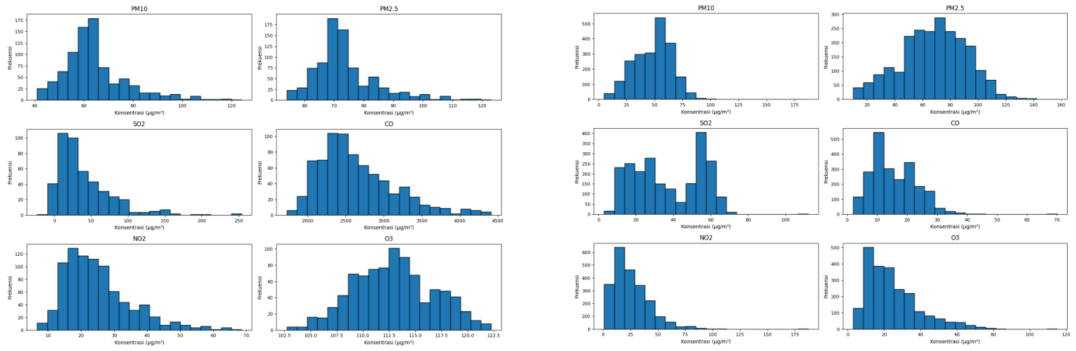
Berikut dilampirkan pola sebaran dan distribusi nilai dari bahan pencemar di DKI Jakarta yang mencakup CO, NO_2, SO_2, O_3, dan AOD menggunakan peta heatmap dan histogram:



Gambar 1. Pola Sebaran Polutan Tahun 2022-2024

Peta heatmap di atas dibuat menggunakan *platform* GEE dengan visualisasinya mengikuti format pada masing-masing jenis polutan. Di dalam satelit Sentinel-5P umumnya menggunakan palet warna berupa gradasi dari warna hitam, biru, ungu, biru kehijauan, hijau, kuning, dan merah. Secara berurutan warna tersebut mengilustrasikan tingkat kualitas udara dari baik hingga buruk. Berdasarkan Gambar 1., dapat diketahui bahwa pola sebaran kadar CO, NO_2, SO_2, O_3, dan AOD di wilayah DKI Jakarta mengalami fluktuasi sepanjang tahun 2022 hingga 2024. Pola sebaran CO menunjukkan bahwa wilayah Jakarta Barat dan sebagian Jakarta Utara memiliki konsentrasi yang tinggi, ditandai dengan warna oranye dan merah. Untuk polutan NO_2, konsentrasi tertinggi terlihat pada tahun 2022 karena intensitas warna merah yang cukup merata pada Jakarta Selatan dan Jakarta Pusat. Hal serupa juga terjadi di tahun 2023 hingga

2024 terlihat bahwa polutan NO_2 masih tinggi di dua daerah tersebut walaupun terlihat terdapat sedikit penurunan. Untuk polutan SO_2 menunjukkan kadar yang tinggi di tahun 2023 dan 2024, khususnya di wilayah Jakarta Timur dan Jakarta Selatan. Untuk wilayah lainnya terlihat lebih bervariasi dari warna biru, hijau, hingga kekuningan. Untuk polutan O_3, pada tahun 2022 hingga 2024 terlihat konsentrasi tertinggi ada di wilayah Jakarta Barat, Jakarta Pusat, dan Jakarta Utara, ditandai dengan warna merah yang pekat dan merata di beberapa bagian utara peta Jakarta. Sementara itu, untuk wilayah Jakarta Selatan dan Jakarta Timur polutan O_3 memiliki konentrasi yang lebih rendah ditandai dengan gradasi warna biru dan hijau. Untuk nilai AOD menunjukkan kecenderungan kadar tinggi pada tahun 2022. Namun, hanya pada sebagian kecil di wilayah Jakarta Utara. Sementara di wilayah lainnya cenderung memiliki konsentrasi yang lebih rendah ditandai dengan gradasi warna biru dan hijau.



(a) Histogram Data Citra Satelit $\mu g/m^3$ (b) Histogram Data SPKU

Gambar 2. Histogram Data Polutan

Grafik di atas menunjukkan distribusi nilai setiap polutan yang diperoleh dari data SPKU dan citra satelit yang telah dikonversi ke dalam satuan µg/m³. Berdasarkan Gambar 2., kadar polutan PM10, PM2.5, SO_2, CO, dan NO_2 cenderung memiliki distribusi yang miring ke kanan, yang mengindikasikan bahwa sebagian besar nilai berada di bawah rata-rata atau tergolong rendah. Sedangkan pada polutan SO_2, terlihat adanya nilai yang sangat tinggi namun jumlahnya sedikit, yang menunjukkan keberadaan pencilan. Sementara itu, distribusi polutan O_3 tampak simetris, menandakan bahwa sebagian besar nilainya berada di sekitar rata-rata. Namun, perlu diketahui bahwa meskipun nilai polutan mayoritas rendah, hal tersebut belum dapat disimpulkan sebagai kualitas udara yang baik. Hal ini disebabkan karena data yang digunakan masih dalam satuan µg/m³ dan belum dikonversi ke dalam nilai ISPU sehingga nilai konsentrasi polutan belum secara langsung merepresentasikan kategori kualitas udara. Distribusi nilai polutan berdasarkan data SPKU menunjukkan bahwa polutan CO, NO_2, dan O_3 cenderung memiliki distribusi yang miring ke kanan, yang mengindikasikan bahwa sebagian besar nilai berada di bawah rata-rata.

Sementara itu, distribusi polutan PM10, PM2.5, dan SO_2 tampak lebih simetris, meskipun terdapat sejumlah nilai tinggi dengan frekuensi yang rendah. Dapat diketahui juga bahwa sebagian besar nilai ISPU dari data SPKU berada pada rentang 40 hingga 70, yang mengindikasikan bahwa pada data SPKU klasifikasi kualitas udara paling dominan berada pada kategori sedang.

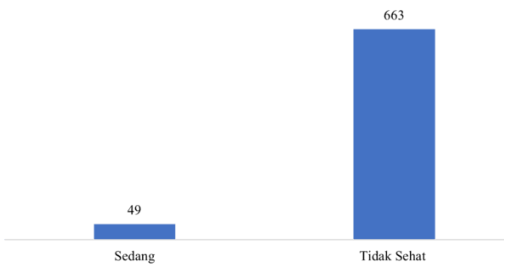
3.2. Penanganan Data Hilang dan Ketidakseimbangan Data

Tabel 3. Persentase data hilang pada data citra satelit dan data SPKU

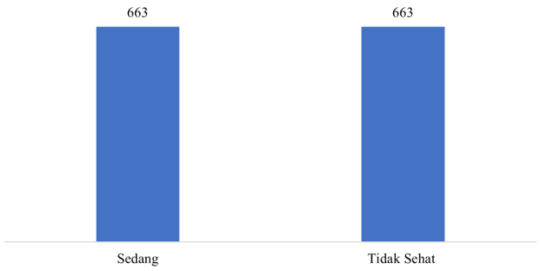
Variabel	Persentase data hilang	
	Data Citra Satelit	Data SPKU
PM10	6.17	5.38
PM2.5	6.17	0.88
Sulfur Dioxide (SO_2)	50.17	1.06
Carbon Monoxide (CO)	15.82	1.01
Nitrogen Dioxide (NO_2)	4.71	1.85
Ozone (O_3)	3.59	0.79

Pada Tabel 3, menunjukkan persentase data hilang untuk masing-masing variabel polutan pada dua sumber data, yaitu data citra satelit dan data SPKU (Stasiun Pemantauan Kualitas Udara). Berdasarkan tabel tersebut diketahui bahwa polutan dengan masalah data hilang tertinggi adalah *Sulfur Dioxide* (SO_2) sebesar 50,17% yang menunjukkan masalah signifikan dalam ketersediaan data SO_2 dari citra satelit. Diikuti oleh data *Carbon Monoxide* (CO) sebesar 15,82%, PM10 dan PM2.5 masing-masing kehilangan 6,17%, dan NO_2 dan O_3 relatif lebih kecil, yaitu 4,71% dan 3,59%. Hal ini mengindikasikan bahwa akurasi dan kelengkapan data dari citra satelit bervariasi antar polutan, dan perlunya strategi imputasi yang tepat dalam penanganan data hilang. Sementara itu, secara umum data SPKU memiliki tingkat kehilangan data yang jauh lebih rendah dibandingkan data citra satelit. Persentase tertinggi terjadi data hilang ada pada PM10 sebesar 5,38%, tetapi angka ini masih tergolong rendah. Polutan lainnya juga memiliki presentase data hilang yang lebih rendah seperti PM2.5 sebesar 0,88%, SO_2 sebesar 1,06%, CO sebesar 1,01%, NO_2 sebesar 1,85%, dan O_3 sebesar 0,79%. Ini menunjukkan bahwa ketersediaan data pada data SPKU jauh lebih lengkap dibandingkan data citra satelit. Dengan menggunakan median imputation memastikan bahwa setiap polutan yang memiliki masalah data hilang akan dapat tertangani sehingga dapat menghindari terjadinya bias saat melakukan klasifikasi. Meskipun demikian, metode imputasi merupakan bagian dari pendekatan pemodelan yang dapat memengaruhi hasil klasifikasi. Hali itu disebabkan karena penggantian nilai hilang dengan median berpotensi mengurangi variasi data dan meredam nilai ekstrem. Selain itu, tingginya proporsi data hilang pada beberapa polutan citra satelit

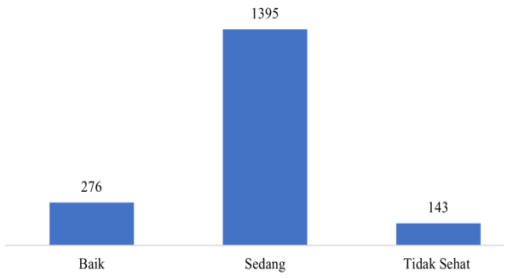
mencerminkan keterbatasan sensor dan kondisi observasi, seperti adanya tutupan awan. Oleh karena itu, hasil klasifikasi yang diperoleh dalam penelitian ini, perlu diinterpretasikan secara hati-hati dan tidak dimaknai sebagai representasi kondisi absolut kualitas udara, tetapi sebagai estimasi kondisi kualitas udara berbasis data.



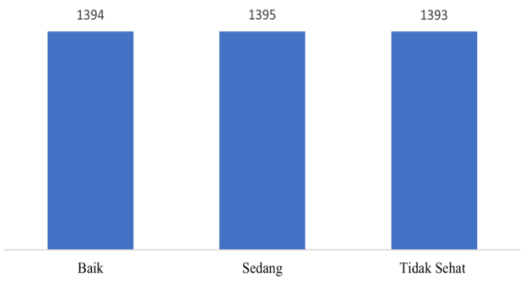
(a) Distribusi Kelas Sebelum Resampling pada Data Citra Satelit



(b) Distribusi Kelas Seteah Resampling pada Data Citra Satelit



(c) Distribusi Kelas Sebelum Resampling pada Data SPKU



(d) Distribusi Kelas Setelah Resampling pada Data SPKU

Gambar 3. Distribusi Kelas Data

Berdasarkan Gambar 3., distribusi kelas pada data citra satelit dan data SPKU sebelum dan sesudah dilakukan *resampling* dengan metode SMOTE Tomek Links menunjukkan perubahan yang signifikan. Sebelum dilakukan *resampling*, data citra satelit sangat tidak seimbang, di mana kelas "Sedang" hanya memiliki 49 sampel, sedangkan kelas "Tidak Sehat" sebanyak 663 sampel. Ketidakseimbangan ini berisiko menurunkan performa model klasifikasi karena algoritma cenderung bias terhadap kelas mayoritas. Setelah diterapkan SMOTE Tomek Links, jumlah sampel dari kedua kelas menjadi seimbang, masing-masing 663 sampel, yang dihasilkan dari proses *oversampling* terhadap kelas minoritas oleh SMOTE dan pembersihan data ambang oleh Tomek Links. Hal serupa terjadi pada data SPKU, yang awalnya didominasi oleh kelas "Sedang" dengan 1395 sampel, sementara kelas "Baik" dan "Tidak Sehat" masing-masing hanya 276 dan 143 sampel. Setelah resampling, ketiga kelas menjadi hampir seimbang, masing-masing sekitar 1394–1395 sampel. Hasil ini menunjukkan bahwa SMOTE Tomek Links sangat

efektif dalam menangani ketidakseimbangan kelas, sehingga dapat meningkatkan kemampuan model untuk mengenali semua kelas secara adil dan menghindari bias terhadap kelas mayoritas.

3.3. Perbandingan Hasil Klasifikasi Kualitas Udara dengan Data Multisumber

Tabel 4. Parameter SVM terpilih pada data citra satelit dan data SPKU

Parameter	Nilai Parameter	Parameter Terpilih	
		Data Citra Satelit	Data SPKU
C	0,1, 1, 10	10	10
Gamma	Scale, Auto, 0,01, 0,1, 1	Scale	0,1
Kernel	RBF, Linear, Polynomial	Linear	RBF

Berdasarkan Tabel 4, dilakukan pemilihan parameter optimal (*hyperparameter tuning*) untuk algoritma Support Vector Machine (SVM) pada dua jenis data yaitu data citra satelit dan data SPKU. Terdapat tiga parameter utama yang diuji, yaitu C, gamma, dan kernel. Parameter C merupakan nilai regulasi yang mengontrol *trade-off* antara memaksimalkan margin pemisah dan meminimalkan kesalahan klasifikasi. Nilai C = 10 yang terpilih pada kedua jenis data menunjukkan bahwa model lebih menekankan pada minimalisasi kesalahan klasifikasi dengan margin yang lebih kecil, sehingga memungkinkan model yang dihasilkan akan lebih kompleks dan lebih peka terhadap keslaahan klasifikasi. Untuk parameter gamma, yang berperan dalam menentukan pengaruh dari satu data terhadap keputusan batas (khususnya untuk kernel non-linear), model citra satelit memilih nilai scale, sedangkan pada data SPKU, nilai optimal adalah 0,1. Pemilihan gamma "scale" secara otomatis menyesuaikan nilai gamma terhadap jumlah fitur, dan cocok untuk model linear yang digunakan pada data citra satelit. Sementara itu, nilai 0,1 yang digunakan pada data SPKU bersifat cukup konservatif, memberi pengaruh yang tidak terlalu sempit, tetapi tetap menangkap pola non-linear dengan baik. Perbedaan yang mencolok juga terlihat pada kernel yang digunakan. Untuk data citra satelit, kernel linear dipilih, menandakan bahwa hubungan antar fitur pada data tersebut cukup dapat dipisahkan dengan garis lurus, sehingga dengan model yang lebih sederhana dinilai sudah cukup untuk melakukan klasifikasi. Sebaliknya, pada data SPKU, kernel RBF (Radial Basis Function) dipilih karena mampu menangkap pola non-linear yang lebih kompleks pada data sensor fisik seperti SPKU yang dipengaruhi banyak faktor eksternal.

Secara keseluruhan, pemilihan parameter ini akan sangat berdampak pada performa akhir dari model klasifikasi. Model klasifikasi dengan menggunakan data citra satelit cenderung lebih sederhana, tetapi cukup efektif karena linearitas data, sedangkan model SPKU menggunakan pendekatan non-linear dengan kernel RBF agar dapat mengenali pola-pola yang tidak terlihat

secara linear. Hal ini menunjukkan bahwa karakteristik data sangat memengaruhi strategi pemodelan yang optimal, dan pemilihan parameter yang tepat dapat meningkatkan akurasi, generalisasi, dan stabilitas model dalam memprediksi kelas kualitas udara.

Tabel 5. Hasil evaluasi model SVM pada data citra satelit dan data SPKU dengan dan tanpa resampling

Data	Metrik			
	Accuracy	***Precision***	***Recall***	***F1-Score***
Citra Satelit	99%	100%	96%	98%
Citra Satelit dan SMOTE-Tomek Links	99%	100%	96%	98%
SPKU	98%	98%	94%	96%
SPKU dan SMOTE-Tomek Links	96%	91%	97%	94%

Berdasarkan Tabel 5., hasil evaluasi performa model Support Vector Machine (SVM) ditampilkan untuk dua jenis data yaitu data citra satelit dan data SPKU. Dengan menggunakan dua data yang berbeda, SVM mampu memberikan akurasi yang tinggi yaitu di atas 95%. Hal ini sejalan dengan hasil penelitian sebelumnya yang menunjukkan bahwa algoritma SVM dapat mencapai akurasi 95% hingga 99% dalam klasifikasi kualitas udara, tergantung pada node sensor yang digunakan (Handayani et al., 2020). Pada data citra satelit, model menunjukkan akurasi sangat tinggi yaitu sebesar 99% baik sebelum maupun sesudah *resampling*. *Precision* tetap konsisten di angka 100%, menunjukkan bahwa model sangat baik dalam menghindari prediksi salah untuk kelas positif. Sementara itu, nilai *recall* konsisiten di 96%, yang berarti model mampu mengenali sebagian besar kasus aktual dengan baik. *F1-score* sebesar 98% mengindikasikan keseimbangan yang sangat baik antara *precision* dan *recall*. Ini menandakan bahwa meskipun terdapat ketidakseimbangan kelas awal seperti yang terlihat pada grafik distribusi sebelumnya, model SVM tetap dapat menangani data citra satelit dengan sangat efektif bahkan sebelum dilakukan resampling. Namun, perlu diketahui bahwa akurasi yang tinggi tidak selalu mencerminkan performa algoritma yang baik dalam semua kondisi. Dalam kasus seperti dataset dengan distribusi kelas yang tidak seimbang atau jumlah data yang terbatas, metrik akurasi dapat memberikan gambaran yang menyesatkan dan tidak selalu mencerminkan ketepatan model secara menyeluruh (Chandra et al., 2023).

Berbeda halnya dengan evaluasi model klasifikasi menggunakan data SPKU. Model sebelum *resampling* memperoleh akurasi sebesar 98%, *precision* sebesar 98%, *recall* sebesar 94%, dan *F1-score* sebesar 96%. Setelah dilakukan *resampling* dengan SMOTE-Tomek Links, terjadi sedikit penurunan pada *precision* menjadi 91%. Namun, terdapat peningkatan *recall* menjadi 97%, dengan *F1-score* menurun sedikit menjadi 94%. Penurunan *precision* menunjukkan

bahwa setelah *resampling*, model menjadi lebih sensitif terhadap kelas minoritas, tetapi dengan sedikit peningkatan kemungkinan kesalahan dalam prediksi kelas positif. Namun, kenaikan *recall* memperlihatkan bahwa model menjadi lebih baik dalam mendeteksi semua kelas, terutama kelas-kelas yang sebelumnya minoritas seperti "Baik" dan "Tidak Sehat" sehingga kemampuan generalisasi terhadap kelas-kelas minoritas menjadi meningkat. Hal ini sejalan dengan penelitian sebelumnya yang menyebutkan bahwa meskipun penurunan nilai *precision* secara numerik umumnya lebih besar dibandingkan peningkatan *recall*, namun dalam permasalahan klasifikasi yang tidak seimbang, *recall* biasanya lebih penting karena kesalahan negatif (*false negative*) seringkali memiliki dampak yang lebih besar dibandingkan kesalahan positif (*false positive*) (Boardman, Biron, & Rimbey, 2018). Ini menunjukkan bahwa penggunaan SMOTE-Tomek Links bermanfaat untuk meningkatkan sensitivitas model terhadap ketidakseimbangan data.

Secara keseluruhan, penggunaan SMOTE-Tomek Links pada data citra satelit tidak memberikan perubahan signifikan karena performa awal model sudah sangat tinggi dan data cenderung linier. Namun, untuk data SPKU, penerapan SMOTE-Tomek Links terbukti meningkatkan kemampuan model dalam mengenali kelas minoritas, meskipun dengan sedikit *trade-off* pada *precision*. Oleh karena itu, penggunaan SMOTE-Tomek Links lebih direkomendasikan pada model klasifikasi yang menggunakan data SPKU karena memberikan manfaat dalam memperbaiki ketidakseimbangan kelas, sedangkan untuk data citra satelit penggunaannya tidak terlalu signifikan meningkatkan peforma model klasifikasi.

Tabel 6. Hasil *Precision* untuk Setiap Kategori

Data	***Precision***		
	Baik	**Sedang**	**Tidak Sehat**
Citra Satelit	-	100%	99%
Citra Satelit dan SMOTE-Tomek Links	-	100%	99%
SPKU	97%	98%	100%
SPKU dan SMOTE-Tomek Links	85%	99%	88%

Tabel 7. Hasil *Recall* untuk Setiap Kategori

Data	*Recall*		
	Baik	Sedang	Tidak Sehat
Citra Satelit	-	92%	100%
Citra Satelit dan SMOTE-Tomek Links	-	92%	100%
SPKU	94%	99%	89%
SPKU dan SMOTE-Tomek Links	99%	95%	97%

Tabel 8. Hasil *F1-Score* untuk Setiap Kategori

Data	*F1-Score*		
	Baik	Sedang	Tidak Sehat
Citra Satelit	-	96%	100%
Citra Satelit dan SMOTE-Tomek Links	-	96%	100%
SPKU	96%	99%	94%
SPKU dan SMOTE-Tomek Links	91%	97%	92%

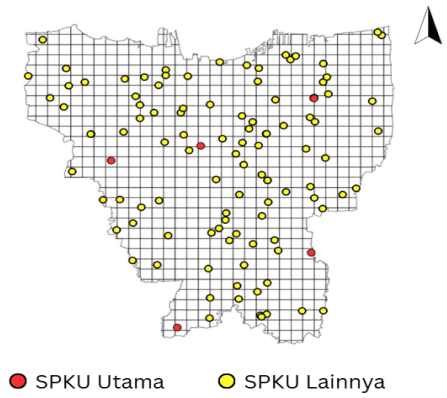
Berdasarkan Tabel di atas metrik evaluasi diperinci berdasarkan tiap kelas (Baik, Sedang, dan Tidak Sehat) untuk model klasifikasi kualitas udara menggunakan algoritma SVM, baik pada data citra satelit maupun data SPKU. Hasil menunjukkan bahwa model SVM pada data citra satelit hanya mengklasifikasikan dua kelas, yaitu "Sedang" dan "Tidak Sehat", tanpa mencakup kelas "Baik". *Precision* model ini sangat tinggi, yaitu 100% untuk kelas Sedang dan 99% untuk Tidak Sehat, serta *recall* masing-masing 92% dan 100%, menghasilkan *F1-score* sebesar 96% dan 100%. Nilai ini mencerminkan model yang sangat presisi dan andal dalam mengidentifikasi kedua kelas tersebut, bahkan hasilnya tetap konsisiten setelah diterapkan SMOTE-Tomek Links. Hal ini mengindikasikan bahwa model dengan data citra satelit sudah optimal sejak awal dan penambahan SMOTE-Tomek tidak berdampak signifikan terhadap performa setiap kelas.

Sebaliknya, pada data SPKU, hasil evaluasi menunjukkan pengaruh yang signifikan dari penggunaan SMOTE-Tomek Links. Sebelum *resampling*, model memiliki *precision* tinggi untuk semua kelas, yaitu 97% (Baik), 98% (Sedang), dan 100% (Tidak Sehat), namun recall pada kelas "Tidak Sehat" masih tergolong rendah (89%), yang menunjukkan masih ada data dengan kelas "Tidak Sehat" yang gagal terklasifikasi dengan benar. Setelah dilakukan SMOTE-Tomek Links, *recall* meningkat secara signifikan menjadi 97%, walaupun *precision* untuk kelas "Baik" turun menjadi 85%. *F1-score* pun menjadi lebih merata, yaitu 91% (Baik), 97% (Sedang), dan 92% (Tidak Sehat). Ini menunjukkan bahwa meskipun ada sedikit penurunan *precision* pada kelas tertentu, secara keseluruhan model menjadi lebih seimbang dalam mengenali seluruh kelas, terutama kelas minoritas yang sebelumnya sulit dikenali oleh model

seperti "Baik" dan "Tidak Sehat". Hal ini mendukung kesimpulan sebelumnya bahwa penggunaan SMOTE-Tomek Links lebih berdampak positif pada data SPKU dibandingkan data citra satelit. Oleh karena itu, strategi resampling lebih relevan dan efektif diterapkan pada dataset dengan jumlah kelas lebih banyak dan distribusi yang tidak seimbang seperti pada data SPKU.

Prediksi kualitas udara yang akurat berpotensi membantu pemerintah dalam mengantisipasi kondisi polusi sebelum mencapai tingkat yang lebih berbahaya. Selain itu, pendekatan berbasis machine learning dapat mendukung pemantauan kualitas udara secara spasial-temporal serta menyediakan informasi tambahan dalam pengambilan kebijakan lingkungan berbasis data. Dengan demikian, model yang dihasilkan dalam penelitian ini menunjukkan peluang untuk dimanfaatkan sebagai salah satu komponen pendukung dalam sistem monitoring lingkungan terkait pemantauan kualitas udara di Jakarta, terutama jika dikombinasikan dengan sistem pemantauan dan kebijakan yang sudah berjalan.

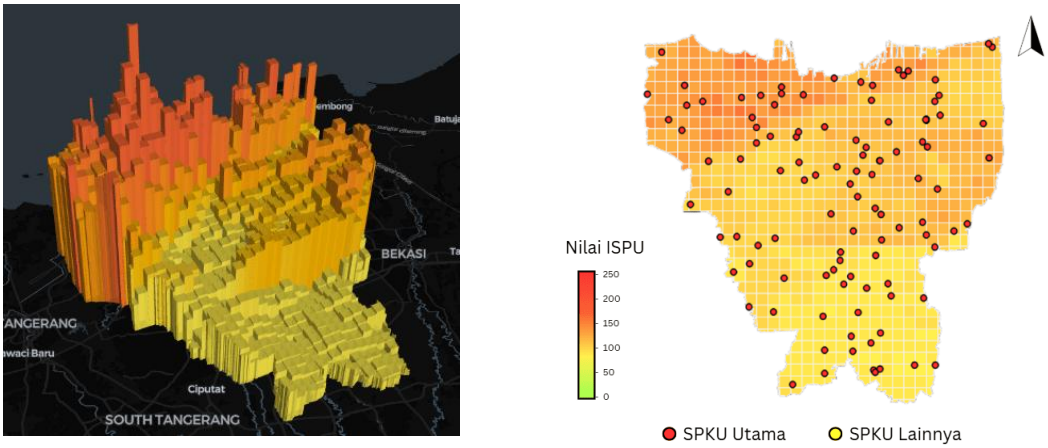
3.4. Pemetaan Spasial Titik SPKU dan Tingkat Kualitas Udara di DKI Jakarta



Gambar 4. Persebaran Titik SPKU di DKI Jakarta

Berdasarkan Gambar 4., DKI Jakarta memiliki total sebanyak 111 titik SPKU yang tersebar pada tiap wilayahnya. Sebelumnya, terdapat 5 SPKU pada tiap-tiap kota yang ditandai dengan titik merah untuk mengambil data di tiap wilayah. Persebaran SPKU cukup merata pada tiap kota meskipun memiliki beberapa kepadatan yang berbeda. Jakarta Pusat dan Jakarta Barat memiliki titik SPKU yang tersebar cukup merata. Perbatasan Jakarta Selatan dan Jakarta Timur memiliki titik SPKU yang sangat padat, berbeda dengan wilayah utara dari Jakarta Selatan yang tidak memiliki titik SPKU. Selain itu, terdapat beberapa wilayah pada Jakarta Utara yang tidak memiliki jangkauan SPKU sama sekali. Persebaran titik SPKU tentunya berpengaruh pada wilayah-wilayah di sekitarnya. Wilayah dengan kepadatan SPKU yang tinggi tentunya akan memudahkan dalam proses pemantauan kualitas udara. Sebaliknya, wilayah-wilayah yang

berada jauh dan tidak berada dalam jangkauan SPKU akan lebih sulit untuk dipantau. Pemantauan melalui data citra satelit menjadi solusi pada penelitian ini untuk mengatasi pemantauan kualitas udara pada wilayah yang tidak terjangkau oleh SPKU.



(a) Visualisasi 3D Tingkat Kualitas Udara di DKI Jakarta Menggunakan Kepler.gl

(b) Sebaran Lokasi SPKU dan Nilai ISPU di DKI Jakarta

Gambar 5. Peta Spasial Tingkat Kualitas Udara DKI Jakarta Pada Bulan Mei 2025

Berdasarkan Gambar 5., Lokasi SPKU di DKI Jakarta tampaknya memiliki korelasi dengan kondisi ISPU di tiap wilayahnya. Pada bulan Mei 2025, seluruh wilayah DKI Jakarta tercatat memiliki kualitas udara yang tergolong tidak sehat berdasarkan kategori ISPU dengan nilai berkisar antara 108,908–138,814. Meski demikian, visualisasi peta menunjukkan gradasi keparahan dalam kategori "tidak sehat", yang divisualisasikan melalui warna (hijau muda, kuning, oranye) serta tinggi bar chart pada peta 3D. Wilayah Jakarta Utara memperlihatkan tingkat pencemaran yang relatif lebih buruk dibandingkan wilayah lainnya, tergambar dari warna oranye dan ketinggian bar chart yang lebih dominan. Sementara itu, Jakarta Selatan cenderung memiliki nilai ISPU yang sedikit lebih rendah dalam kategori "tidak sehat", ditunjukkan oleh warna yang lebih cerah (hijau muda hingga kuning) serta bar chart yang lebih pendek. Secara umum, kondisi udara di DKI Jakarta tidak sehat secara merata, namun terdapat perbedaan tingkat keparahan antarwilayah. Peta spasial ini membantu memvisualisasikan wilayah mana yang mengalami pencemaran lebih berat dibandingkan wilayah lainnya, meskipun semuanya masih dalam rentang kualitas udara tidak sehat. Pada visualisasi tersebut juga terlihat adanya hubungan spasial antara satu grid dengan grid lain, nilai ISPU turun secara bertahap sehingga tidak menyebabkan adanya pencilan pada wilayah-wilayah tertentu.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bagian sebelumnya, terdapat beberapa hal penting yang dapat ditarik sebagai kesimpulan. Berdasarkan analisis deskriptif, tingkat pencemaran udara di DKI Jakarta menunjukkan pola fluktuatif sepanjang tahun 2022–2024. Hal ini dapat dilihat melalui sebaran spasial nilai ISPU yang menunjukkan gradasi warna oranye menuju kuning pada bulan Mei 2025. Sementara distribusi nilai setiap polutan mayoritas miring ke kanan walaupun pada beberapa polutan masih ada yang terindikasi memiliki pencilan. Data citra satelit memiliki proporsi data hilang yang lebih tinggi dibandingkan data SPKU. Penelitian ini mengidentifikasi adanya kelemahan integrasi data akibat ketidaksamaan (*mismatch*) resolusi spasial antara citra satelit yang berbasis grid luas dengan data SPKU yang berbasis titik (*point-based*), yang berkontribusi pada tingginya proporsi data hilang pada citra satelit. Meskipun demikian, evaluasi performa menunjukkan bahwa model berbasis citra satelit tetap optimal sebelum dan sesudah dilakukan *resampling*. Sebaliknya, pada data SPKU, penerapan SMOTE-Tomek Links signifikan berpengaruh terhadap peningkatan *recall* dan keseimbangan antar kelas yang lebih merata. Peta hasil data citra satelit memiliki nilai yang lebih detail untuk tiap-tiap grid lokasi di DKI Jakarta, seluruh wilayah merepresentasikan nilai berdasarkan data hasil olahan. Beberapa wilayah yang tidak terjangkau oleh SPKU memiliki nilai ISPU yang cenderung buruk, sedangkan wilayah dengan jumlah titik SPKU lebih banyak cenderung memiliki kualitas udara lebih baik. Hasil akhirnya menunjukkan bahwa lokasi titik SPKU di tiap wilayah memiliki keterwakilan yang kuat terhadap kondisi kualitas udara di sekitarnya, di mana kerapatan titik pantau membantu memvalidasi gradasi nilai ISPU yang divisualisasikan tanpa adanya pencilan yang ekstrem.

DAFTAR PUSTAKA

Ben-Hur, A. & Weston, J., 2010. A user's guide to support vector machines. In: O. Carugo & F. Eisenhaber, eds. *Data mining techniques for the life sciences*. Totowa, NJ: Humana Press, pp. 223–239.

Boardman, J., Biron, K. & Rimbey, R., 2018. Mitigating the effects of class imbalance using SMOTE and Tomek link undersampling in SAS. In: *Proceedings of the SAS Global Forum 2018 Conference*, Paper 3604-2018.

Cortes, C. & Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), pp. 273–297.

Chandra, W., Suprihatin, B. & Resti, Y., 2023. Median-KNN regressor–SMOTE–Tomek links for handling missing and imbalanced data in air quality prediction. *Symmetry*, 15(4), p. 857. Available at: https://doi.org/10.3390/sym15040857

Gorelick, N. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, pp. 18–27.

Handayani, R., Salamah, R. & Wibowo, A., 2020. Penerapan Support Vector Machine untuk klasifikasi kualitas udara berdasarkan data sensor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(2), pp. 345–352.

Han, J., Zhang, W., Liu, H. & Xiong, H., 2018. Machine learning for urban air quality analytics: A survey. *Journal of the ACM*, 37(4).

Hovi, H.S.W., Hadiana, A.I. & Umbara, F.R., 2022. Prediksi penyakit diabetes menggunakan algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, 4(1), pp. 40–45. Available at: https://doi.org/10.36423/index.v4i1.895

IQAir, 2024. *Indonesia Air Quality Index (AQI) and air pollution information*. Available at: https://www.iqair.com/indonesia (Accessed: 8 June 2025).

IQAir, 2025. *World Air Quality Report 2024: Region and city $PM_{2.5}$ ranking*. Available at: https://www.greenpeace.org/static/planet4-chilstateless/2025/03/edf90b7a-2024_world_air_quality_report_vf.pdf (Accessed: 8 June 2025).

Jayadi, M.R., Pratama, A.Y. & Wibowo, D.A., 2023. Perbandingan algoritma SVM dan KNN dalam klasifikasi kualitas udara di DKI Jakarta. *Jurnal Teknologi dan Sistem Komputer*, 11(1), pp. 20–27.

Kementerian Lingkungan Hidup dan Kehutanan, 2025. *Indeks Standar Pencemaran Udara (ISPU)*. Available at: https://ditppu.menlhk.go.id/portal/read/indeks-standar-pencemar-udara-ispu-sebagai-informasi-mutu-udara-ambien-di-indonesia (Accessed: 8 June 2025).

Kurniawan, D., Sulistiyanti, S.R. & Murdika, U., 2023. Sistem pemantau gas karbon monoksida (CO) dan karbon dioksida (CO_2) menggunakan sensor MQ7 dan MQ-135 terintegrasi Telegram. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(2), pp. 200–206. Available at: https://doi.org/10.23960/jitet.v11i2.2963

Mohammad, L. et al., 2025. Role of MODIS and Sentinel-5P in monitoring air pollution and its impact on vegetation health. *Advances in Space Research*.

Nguyen, T.T. et al., 2015. Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study. *Environmental Research Letters*, 10, 095016.

Nugraha, A.P. & Sasongko, R.A., 2022. Implementasi Grid Search pada optimasi hyperparameter dalam algoritma machine learning. *Jurnal Ilmiah Informatika*, 8(1), pp. 45–52.

Pontoh, R.S. et al., 2023. Air quality mapping in Bandung City. *Atmosphere*, 14, 1444.

Rowley, P. & Karakuş, C., 2022. A multimodal AI approach to high-resolution air quality mapping using Sentinel-5P and ground-based data in Europe. *Environmental Modelling & Software*, 157, 105513. Available at: https://doi.org/10.1016/j.envsoft.2022.105513

Rui-jun, Y., Dan-feng, D. & Feng, Y., 2019. Application of improved KNN algorithm in air quality assessment. In: *Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference*, pp. 108–112.

Savenets, M., 2021. Air pollution in Ukraine: A view from the Sentinel-5P satellite. *Idojaras*, 125(2), pp. 271–290.

World Health Organization, 2024. *Ambient (outdoor) air pollution*. Available at: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health