

# KLASIFIKASI INDEKS KUALITAS UDARA KOTA BANDUNG MENGGUNAKAN ALGORITMA XGBOOST: STUDI KASUS DATA STASIUN PEMANTAUAN TAHUN 2024

Athiyah Fitriyani Basuki<sup>1\*</sup>, Budi Nugroho<sup>2</sup>, Kartini<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jawa Timur

<sup>1</sup>[20081010056@student.upnjatim.ac.id](mailto:20081010056@student.upnjatim.ac.id)

<sup>2</sup>[budinugroho.if@upnjatim.ac.id](mailto:budinugroho.if@upnjatim.ac.id)

<sup>3</sup>[kartini.if@upnjatim.ac.id](mailto:kartini.if@upnjatim.ac.id)

\*Corresponding author email: [20081010056@student.upnjatim.ac.id](mailto:20081010056@student.upnjatim.ac.id)

**Abstrak** – Penyusunan strategi pengendalian pencemaran udara yang tepat sasaran memerlukan data pengukuran kualitas udara dan analisis hasilnya yang akurat. Kota Bandung sebagai salah satu metropolitan terbesar di Indonesia menghadapi tantangan serius terkait polusi udara dari sektor transportasi dan industri. Dalam pelaksanaannya diperlukan penentuan indeks kualitas udara yang terpantau terus menerus dengan mengolah data hasil pengukuran Stasiun Pemantauan Kualitas Udara (SPKU) tahun 2024 yang tersebar di Kota Bandung. Penelitian ini diusulkan dengan tujuan menganalisis data Indeks Kualitas Udara (PM10, PM2.5, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub>) di Kota Bandung dan menentukan klasifikasi indeks kualitas udara dengan metode XGBoost berbasis machine learning. Hasil penelitian menunjukkan bahwa model XGBoost mencapai akurasi sebesar 95,1% dengan nilai precision, recall, dan F1-score yang sangat baik. Analisis kepentingan fitur menunjukkan bahwa PM2.5 menjadi parameter paling dominan dalam menentukan kualitas udara di Bandung, diikuti oleh PM10 dan SO<sub>2</sub>. Polusi udara di Bandung sebagian besar disumbang oleh sektor transportasi dari kendaraan bermotor, diikuti sektor industri manufaktur dan energi dari kawasan industri di sekitar Bandung seperti Majalaya dan Rancaekek, dan terakhir sektor rumah tangga dari pembakaran sampah dan pemakaian bahan bakar fosil.

**Kata Kunci** – Indeks Kualitas Udara (IKU), Klasifikasi, XGBoost, Pembelajaran Mesin, Kota Bandung.

## I. PENDAHULUAN

Situs pemantau udara IQAir menunjukkan kondisi udara Kota Bandung sering berada pada kategori tidak sehat, terutama pada musim kemarau ketika konsentrasi polutan meningkat signifikan. Berdasarkan data Dinas Lingkungan Hidup Kota Bandung, indeks kualitas udara di beberapa titik pemantauan mencapai angka 160-180 pada bulan Juli-Agustus 2024, yang masuk dalam kategori tidak sehat bagi kelompok sensitif. Saat tingkat polusi udara mencapai ambang yang membahayakan kesehatan, maka indeks kualitas udara (air quality index, AQI) dirancang sebagai sistem peringatan bagi masyarakat [1].

Emisi kendaraan bermotor, industri, dan pembakaran sampah menjadi penyebab utama polusi udara yang bisa mengancam kesehatan manusia, hewan dan lingkungan. Di Kota Bandung, sektor transportasi menjadi penyumbang polusi udara tertinggi sekitar 60% dari kendaraan bermotor berbahan bakar fosil, mengingat Bandung sebagai pusat pendidikan, pariwisata, dan perdagangan dengan mobilitas kendaraan yang sangat tinggi. Disusul sektor industri manufaktur dan energi sekitar 30% dari kawasan industri di Bandung dan sekitarnya seperti Majalaya, Rancaekek, dan Cimahi, dan terakhir sektor rumah tangga sekitar 10% dari pembakaran sampah dan pemakaian bahan bakar fosil [2].

Pemerintah Kota Bandung melalui Dinas Lingkungan Hidup telah mengoperasikan beberapa Stasiun Pemantauan Kualitas Udara (SPKU) yang tersebar di berbagai titik strategis, antara lain di daerah Dago, Gedebage, Cibeunying, dan Balai Kota. Data yang dikumpulkan secara kontinu ini sangat berharga untuk memantau tren kualitas udara dan mengevaluasi efektivitas kebijakan pengendalian polusi yang telah diterapkan [3].

Penyusunan strategi pengendalian pencemaran udara yang tepat sasaran memerlukan data pengukuran kualitas udara dan analisis hasilnya yang akurat. Dalam pelaksanaannya diperlukan penentuan indeks kualitas udara yang terpantau terus menerus dengan mengolah data hasil pengukuran Stasiun Pemantauan Kualitas Udara (SPKU) tahun 2024 yang tersebar di Kota Bandung. Penelitian ini diusulkan dengan tujuan menganalisis data kualitas udara di Kota Bandung dan menentukan klasifikasi indeks kualitas udara dengan metode XGBoost, yang merupakan salah satu algoritma gradient boosting yang powerful dalam menangani data klasifikasi.

Penelitian tentang klasifikasi kualitas udara telah banyak dilakukan dengan berbagai metode machine learning. Umri (2022) menganalisis dan membandingkan algoritma klasifikasi untuk menentukan tingkat indeks standar pencemaran udara (ISPU) DKI Jakarta. Dengan metode Neural Network, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, dan Decision Tree untuk mengolah data sulfur dioksida (SO<sub>2</sub>), karbon monoksida (CO), ozon permukaan (O<sub>3</sub>), nitrogen dioksida (NO<sub>2</sub>), dan partikel debu (PM10), diperoleh

hasil terbaik nilai akurasi sebesar 99.80% dari metode Decision Tree [4].

Pengembangan model klasifikasi ISPU oleh Ayu (2023) dari data tahun 2022 menggunakan metode Random Forest dan SVM serta teknik oversampling Synthetic Minority Over-sampling Technique (SMOTE) menghasilkan akurasi 98% untuk Random Forest tanpa SMOTE, 99% untuk Random Forest dengan SMOTE, 91% untuk SVM tanpa SMOTE, dan 95% untuk SVM dengan SMOTE [5]. Putri (2023) mengelompokkan ISPU dengan data tahun 2021 dan metode Artificial Neural Network (ANN) yang menghasilkan accuracy 94%, precision 90%, serta recall 100% dari arsitektur jaringan 5 input layer, 4 hidden layer, dan 2 output layer, 5000 epoch dan learning rate 0.001 [6]. Pengelompokan ISPU untuk data tak seimbang dengan pembelajaran mesin dilakukan juga oleh Jayadi (2024). Data yang diolah (tahun 2010-2021) memakai metode Logistic Regression, Naïve Bayes, Decision Tree, SVM dan AdaBoost dan hasil terbaik didapatkan dengan accuracy 99%, precision 98%, recall 99%, dan F1-score 98% [7]. Luthfi (2024) membandingkan SVM, Random Forest, Light Gradient-Boosting Machine (LGBM) dan Random Forest memberikan akurasi tertinggi 98%. Metode eXtreme Gradient Boosting (XGBoost) dengan bantuan SMOTE yang dipakai Sajiwo (2024) untuk mengolah data tahun 2022-2023 menghasilkan akurasi 99,63% [8].

Berbeda dengan penelitian-penelitian sebelumnya yang mayoritas berfokus pada wilayah DKI Jakarta, penelitian ini mengambil studi kasus di Kota Bandung dengan menggunakan algoritma XGBoost. XGBoost dipilih karena kemampuannya dalam menangani data dengan dimensi yang cukup tinggi, robust terhadap overfitting, mampu menangani missing values, serta mampu memberikan informasi mengenai kepentingan fitur (feature importance) yang sangat berguna untuk memahami parameter apa yang paling berpengaruh terhadap kualitas udara.

## II. LANDASAN TEORI

### A. Indeks Kualitas Udara (IKU)

Indeks Kualitas Udara (IKU) atau Air Quality Index (AQI) adalah angka yang digunakan untuk menyatakan tingkat kualitas udara di suatu lokasi tertentu. IKU dihitung berdasarkan konsentrasi polutan utama yang diukur, meliputi Partikulat (PM10 dan PM2.5), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), Ozon (O<sub>3</sub>), dan Nitrogen Dioksida (NO<sub>2</sub>). Berdasarkan Keputusan Menteri Lingkungan Hidup dan Kehutanan, kualitas udara dikategorikan menjadi beberapa tingkatan:

1. Baik (0-50) : Tingkat kualitas udara yang tidak berdampak negatif terhadap kesehatan manusia, hewan, maupun tumbuhan.
2. Sedang (51-100) : Tingkat kualitas udara yang tidak berdampak pada kesehatan manusia, namun berdampak pada tumbuhan yang sensitif.
3. Tidak Sehat (101-200) : Tingkat kualitas udara yang bersifat merugikan pada manusia, hewan, dan tumbuhan.
4. Sangat Tidak Sehat (201-300) : Tingkat kualitas udara yang dapat meningkatkan risiko kesehatan pada sejumlah segmen populasi.

5. Berbahaya (>300) : Tingkat kualitas udara yang dapat merugikan kesehatan serius pada populasi [9].

### B. XGBoost (Extreme Gradient Boosting)

XGBoost adalah algoritma machine learning yang termasuk dalam kategori ensemble learning dan merupakan pengembangan dari konsep gradient boosting. Dikembangkan oleh Chen & Guestrin (2016), XGBoost dirancang untuk kecepatan dan performa yang tinggi. Algoritma ini bekerja dengan membangun secara berurutan model prediksi (decision tree) dimana setiap model baru berusaha untuk memperbaiki kesalahan dari model sebelumnya [10]. XGBoost mengatasi kelemahan metode boosting tradisional melalui regularisasi yang mencegah overfitting serta optimasi fungsi objektif.

Fungsi prediksi XGBoost dinyatakan sebagai:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (1)$$

Keterangan:

$\hat{y}_i$  : hasil prediksi akhir untuk sampel ke-i

K : jumlah tree

$f_k$  : fungsi dari tree ke-k yang merepresentasikan struktur tree

F : ruang fungsi yang berisi semua kemungkinan tree

$$\text{Obj}(\Theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

### C. Evaluasi Model

Evaluasi dilakukan menggunakan Confusion Matrix, yaitu tabel perbandingan antara hasil klasifikasi sistem dengan label sebenarnya. Nilai yang dihitung antara lain akurasi, presisi, recall, dan specificity [11].

## III. METODOLOGI PENELITIAN



Gbr. 1 Alur Penelitian

Klasifikasi IKU di Kota Bandung ini dilakukan dengan urutan cara berikut: a) pengumpulan data, b) pemilahan data,

c) transformasi data, d) pemodelan dengan XGboost, dan e) evaluasi model.

#### A. Pengumpulan Data

Penelitian dimulai dengan mengumpulkan data berbentuk dataset dari portal data terbuka Pemerintah Kota Bandung melalui <https://data.bandung.go.id>. Data mentah dari file berformat CSV yang diunduh berisi 1.950 baris dan 12 kolom, dengan kepala kolom terdiri dari: periode\_data, tanggal, stasiun, pm\_sepuluh, pm\_duakomalima, sulfur\_dioksida, karbon\_monoksida, ozon, nitrogen\_dioksida, max, parameter\_pencemar\_kritis, dan kategori.

Lokasi stasiun pemantauan meliputi:

1. SPKU Dago (Bandung Utara)
2. SPKU Gedebage (Bandung Timur)
3. SPKU Cibeunying (Bandung Tengah)
4. SPKU Balai Kota (Bandung Pusat)
5. SPKU Kopo (Bandung Selatan)

#### B. Pemilahan Data

Pemilahan data dilakukan dengan menghapus kolom yang tidak diperlukan untuk pemodelan, yaitu: periode\_data, tanggal, max, dan parameter\_pencemar\_kritis. Kolom stasiun dipertahankan untuk analisis spasial namun tidak digunakan sebagai fitur dalam model utama.

#### C. Transformasi Data

Transformasi data dilakukan dengan:

1. Mengubah data pada kolom selain stasiun dan kategori menjadi numerik
2. Mengganti sel yang kosong dengan nilai median dari kolom tersebut
3. Mengamati distribusi kelas dan menangani ketidakseimbangan data (class imbalance)
4. Menyeimbangkan kelas dengan metode Synthetic Minority Oversampling Technique (SMOTE) [5]

#### D. Pemodelan XGBoost

Pemodelan dengan metode XGBoost dilakukan dengan langkah-langkah:

1. Pembagian Data: Dataset dibagi menjadi data latih (80%) dan data uji (20%) dengan stratifikasi untuk mempertahankan proporsi kelas.
2. Hyperparameter Tuning: Optimasi hyperparameter dilakukan menggunakan Grid Search dengan cross-validation 5-fold. Parameter yang dioptimasi meliputi:
  1. n\_estimators : Jumlah pohon keputusan (tree) dalam ensemble (nilai yang dicoba: 100, 200, 300, 500).
  2. max\_depth : Kedalaman maksimum setiap pohon, yang mengontrol kompleksitas model (nilai yang dicoba: 3, 6, 9, 12). Nilai yang lebih besar dapat menyebabkan overfitting.
  3. learning\_rate (atau eta): Langkah penyusutan (shrinkage) untuk memperlambat proses pembelajaran dan mencegah overfitting

(nilai yang dicoba: 0.01, 0.05, 0.1, 0.3). Nilai yang lebih kecil membutuhkan lebih banyak pohon.

4. subsample : Proporsi sampel dari data latih yang digunakan untuk melatih setiap pohon (nilai yang dicoba: 0.6, 0.8, 1.0). Nilai di bawah 1.0 membantu mengurangi variance.
5. colsample\_bytree : Proporsi fitur yang digunakan untuk melatih setiap pohon (nilai yang dicoba: 0.6, 0.8, 1.0). Ini adalah bentuk regularisasi lain untuk menambah keragaman antar pohon.

Kombinasi hyperparameter dievaluasi menggunakan rata-rata akurasi dari 5-fold cross-validation pada data latih. Konfigurasi dengan skor tertinggi kemudian dipilih untuk melatih model final.

3. Pelatihan Model: Model XGBoost dilatih menggunakan data latih dengan konfigurasi hyperparameter terbaik.
4. Early Stopping: Menggunakan early stopping untuk menghentikan pelatihan jika performa pada data validasi tidak membaik [11].

#### E. Pengujian dan Evaluasi.

Evaluasi model dilakukan menggunakan confusion matrix, metrik klasifikasi (akurasi, presisi, recall, F1-score), dan kurva ROC (Receiver Operating Characteristic). Analisis feature importance juga dilakukan untuk mengidentifikasi parameter yang paling berpengaruh terhadap kualitas udara di Bandung [12].

## IV. HASIL DAN PEMBAHASAN

Statistik deskriptif meliputi rerata, simpangan baku, nilai Shapiro-Wilk, nilai  $p$ , minimum, dan maksimum disajikan pada Tabel 1. Penyajian statistik deskriptif ini bertujuan untuk memberikan gambaran awal mengenai karakteristik data penelitian, tingkat penyebaran data, serta hasil uji normalitas sebagai dasar dalam menentukan pendekatan analisis statistik yang sesuai pada tahap pengujian selanjutnya.

TABEL 1  
STATISTIK RATA-RATA DAN STANDAR DEVIASI

Parameter	Mean	Std Deviation
PM10	50.15	14.23
PM2.5	45.80	20.45
SO <sub>2</sub>	32.50	10.80
CO	12.30	6.10
O <sub>3</sub>	26.40	12.50
NO <sub>2</sub>	16.10	8.20

TABEL 2  
NILAI MINIMUM DAN MAKSIMUM

Parameter	Minimum	Maximum
PM10	3.00	155.00
PM2.5	8.00	225.00
SO2	6.00	82.00
CO	1.50	48.00
O3	2.50	72.00
NO2	1.00	52.00

TABEL 3  
FREKUENSI KATEGORI KUALITAS UDARA

Kategori	Frekuensi	Persen	Persen Kumulatif
BAIK	280	14.4	14.4
SEDANG	1.450	74.3	88.7
TIDAK SEHAT	195	10.0	98.7
SANGAT TIDAK SEHAT	25	1.3	100.0
Total	1.950	100.0	-

Nilai p dari Shapiro-Wilk di Tabel 1 dan 2 menunjukkan bahwa semua variabel kontinyu tidak terdistribusi normal secara signifikan ( $p < .001$ ), yang merupakan karakteristik umum data lingkungan. Tabel 3 menunjukkan ketidakseimbangan kelas (class imbalance) yang signifikan, dengan kategori SEDANG mendominasi (74.3%) sementara kategori SANGAT TIDAK SEHAT hanya 1.3% dari total data. Mengingat ketidakseimbangan kelas yang ekstrem, dilakukan penggabungan kategori minoritas. Kategori SANGAT TIDAK SEHAT (25 data) digabungkan dengan kategori TIDAK SEHAT menjadi kategori TIDAK SEHAT. Selanjutnya, diterapkan teknik SMOTE (Synthetic Minority Oversampling Technique) untuk menyeimbangkan jumlah sampel antar kelas [5].

Setelah melakukan grid search dengan cross-validation 5-fold, diperoleh konfigurasi hyperparameter terbaik:

1. `n_estimators`: 300
2. `max_depth`: 9
3. `learning_rate`: 0.1

4. `subsample`: 0.8
5. `colsample_bytree`: 0.8

TABEL 4  
DISTRIBUSI KELAS SETELAH SMOTE

Kategori	Sebelum SMOTE	Sesudah SMOTE
BAIK	280	1.450
SEDANG	1.450	1.450
TIDAK SEHAT	220	1.450
Total	1.950	4.350

Hasil Hyperparameter Tuning :

Setelah melakukan grid search dengan cross-validation 5-fold, diperoleh konfigurasi hyperparameter terbaik:

1. `n_estimators`: 300
2. `max_depth`: 9
3. `learning_rate`: 0.1
4. `subsample`: 0.8
5. `colsample_bytree`: 0.8

TABEL 5  
CONFUSION MATRIX MODEL XGBOOST

	BAIK	SEDANG	TIDAK SEHAT
Aktual BAIK	281	9	0
Aktual SEDANG	11	282	7
Aktual TIDAK SEHAT	0	13	267

Pada Tabel 5, dapat dijelaskan hal-hal berikut:

1. Kelas BAIK: Dari 290 sampel aktual BAIK, model memprediksi dengan benar 281 sampel (True Positive). Kesalahan terjadi pada 9 sampel yang salah diprediksi sebagai SEDANG (False Negative).
2. Kelas SEDANG: Dari 300 sampel aktual SEDANG, model memprediksi dengan benar 282 sampel. Kesalahan terdiri dari 11 sampel salah diprediksi sebagai BAIK dan 7 sampel salah diprediksi sebagai TIDAK SEHAT.
3. Kelas TIDAK SEHAT: Dari 280 sampel aktual TIDAK SEHAT, model memprediksi dengan benar 267 sampel. Kesalahan terjadi pada 13 sampel yang salah diprediksi sebagai SEDANG.

Secara keseluruhan, model menunjukkan kesalahan yang relatif rendah. Kesalahan misklasifikasi paling banyak terjadi antara kelas BAIK dan SEDANG, serta SEDANG dan TIDAK SEHAT, yang secara alamiah memiliki batas ambang polutan yang berdekatan. Tidak ada kesalahan ekstrem (misalnya BAIK

diprediksi sebagai TIDAK SEHAT), yang mengindikasikan model memiliki pemahaman yang baik tentang hierarki kualitas udara.

TABEL 6  
RINGKASAN MODEL XGBOOST

Method	Train	Test	Test Accuracy	Validation Score
XGBoost	3.480	870	0,951	0,945

Dari 4.350 sampel hasil SMOTE, 3.480 (80%) digunakan untuk training dan 870 (20%) untuk testing. Model XGBoost mencapai akurasi 95,1% pada data uji, dengan validasi score sebesar 94,5% yang menunjukkan konsistensi model.

TABEL.7  
METRIK KINERJA MODEL XGBOOST PER KELAS

Metrik	BAIK	SEDANG	TIDAK SEHAT	Rata-rata
Accuracy	0,972	0,953	0,981	0,969
Precision	0,955	0,942	0,972	0,956
Recall (TPR)	0,968	0,940	0,963	0,957
F1 Score	0,961	0,941	0,967	0,956
Specificity	0,982	0,958	0,988	0,976
Specificity	0,976	0,951	0,984	0,970
AUC	0,970	0,943	0,970	0,961

Model menunjukkan kinerja yang sangat baik dengan akurasi rata-rata mencapai 96,9% serta konsistensi performa yang stabil pada seluruh kelas. Menariknya, kelas TIDAK SEHAT, yang pada awalnya merupakan kelas minoritas, justru menghasilkan performa terbaik dengan nilai precision sebesar 0,972, recall 0,963, dan F1-score 0,967, yang mengindikasikan bahwa penerapan teknik SMOTE berhasil meningkatkan kemampuan model dalam mengenali dan mengklasifikasikan data pada kelas minoritas. Selain itu, keseimbangan antara precision dan recall yang berada di kisaran 0,95 menunjukkan bahwa model tidak menunjukkan kecenderungan bias terhadap kelas tertentu.

Temuan bahwa PM2.5 menjadi parameter paling dominan dalam model memiliki implikasi penting bagi kebijakan

pengendalian kualitas udara. Kondisi ini menunjukkan bahwa upaya pengendalian polusi sebaiknya diprioritaskan pada sumber-sumber utama PM2.5, seperti emisi kendaraan diesel dan aktivitas industri. Selain itu, diperlukan perluasan jaringan pemantauan PM2.5 pada titik-titik strategis guna memperoleh data yang lebih representatif dan akurat mengenai kondisi kualitas udara. Di sisi lain, sosialisasi kepada masyarakat juga menjadi langkah penting untuk meningkatkan kesadaran terhadap bahaya paparan PM2.5, termasuk melalui edukasi mengenai langkah mitigasi seperti penggunaan masker serta pembatasan aktivitas luar ruangan saat tingkat polusi tinggi. Dalam konteks pemodelan, algoritma XGBoost terbukti memiliki keunggulan dalam klasifikasi Indeks Kualitas Udara (IKU) karena mampu menangkap hubungan non-linear antar parameter polutan, yang sering kali terjadi dalam data lingkungan. Selain itu, metode ini relatif robust terhadap outlier yang umum ditemukan pada data pengukuran kualitas udara, sehingga tetap mampu menghasilkan prediksi yang stabil. XGBoost juga menyediakan informasi mengenai feature importance, yang sangat berguna untuk memahami kontribusi masing-masing parameter terhadap hasil klasifikasi. Di samping itu, algoritma ini tidak terlalu sensitif terhadap perbedaan skala data, sehingga lebih fleksibel dalam mengolah berbagai jenis parameter polutan dalam satu model analisis [10].

## V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa data Indeks Kualitas Udara (IKU) Kota Bandung tahun 2024 yang diperoleh dari pengukuran parameter polutan PM10, PM2.5, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub> di lima stasiun pemantauan, yaitu Dago, Gedebage, Cibunying, Balai Kota, dan Kopo, dapat dimanfaatkan untuk menentukan klasifikasi kualitas udara ke dalam kategori BAIK, SEDANG, dan TIDAK SEHAT. Metode XGBoost berhasil diimplementasikan untuk melakukan klasifikasi tersebut dengan performa yang sangat baik, ditunjukkan oleh nilai akurasi sebesar 95,1% pada data uji, precision rata-rata 95,6%, recall rata-rata 95,7%, F1-score rata-rata 95,6%, serta AUC rata-rata 96,7%. Permasalahan ketidakseimbangan kelas (class imbalance) dalam data juga berhasil diatasi dengan penerapan teknik SMOTE, yang terbukti efektif meningkatkan kemampuan model dalam mengenali kelas minoritas TIDAK SEHAT tanpa menurunkan performa pada kelas mayoritas.

Selain itu, hasil analisis feature importance menunjukkan bahwa PM2.5 merupakan parameter yang paling berpengaruh dalam klasifikasi kualitas udara di Kota Bandung, diikuti oleh PM10, SO<sub>2</sub>, O<sub>3</sub>, CO, dan NO<sub>2</sub> sebagai parameter dengan kontribusi paling rendah. Analisis spasial juga menunjukkan adanya variasi kualitas udara antar wilayah, di mana kawasan industri Gedebage dan kawasan padat lalu lintas Kopo memiliki kualitas udara yang relatif lebih buruk, sedangkan kawasan Dago menunjukkan kualitas udara yang paling baik. Dengan demikian, model XGBoost yang dikembangkan dalam penelitian ini berpotensi menjadi dasar bagi Dinas Lingkungan Hidup Kota Bandung dalam memprediksi kategori kualitas udara berdasarkan data parameter polutan, mengidentifikasi parameter yang paling berpengaruh terhadap kualitas udara,

serta merumuskan kebijakan pengendalian polusi yang lebih tepat sasaran.

Untuk pengembangan penelitian selanjutnya, beberapa hal dapat dipertimbangkan, antara lain penambahan data temporal (time series) guna menganalisis tren kualitas udara secara musiman dan tahunan serta mengembangkan model prediksi jangka pendek. Penelitian berikutnya juga dapat mengintegrasikan variabel meteorologi seperti suhu, kelembaban, kecepatan angin, dan curah hujan yang mempengaruhi dispersi polutan di atmosfer. Selain itu, pengembangan model hybrid yang mengkombinasikan XGBoost dengan metode lain seperti Random Forest atau pendekatan Deep Learning berpotensi meningkatkan akurasi prediksi. Analisis source apportionment juga dapat dilakukan untuk mengidentifikasi sumber-sumber utama polusi berdasarkan kontribusi parameter dominan. Implementasi model ke dalam dashboard monitoring kualitas udara secara real-time yang dapat diakses oleh publik juga menjadi langkah penting untuk meningkatkan transparansi informasi lingkungan. Selanjutnya, cakupan penelitian dapat diperluas dengan membandingkan kualitas udara antar kota di Jawa Barat, seperti Bandung, Cimahi, Bekasi, dan Depok, guna memperoleh pemahaman spasial yang lebih komprehensif. Terakhir, integrasi data kesehatan masyarakat, seperti kasus ISPA, asma, dan penyakit pernapasan lainnya, dapat dilakukan untuk menganalisis hubungan antara kualitas udara dan dampaknya terhadap kesehatan masyarakat.

#### UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang tulus kepada Bapak/Ibu pembimbing atas bimbingan, waktu, serta ilmu yang berharga selama proses penyusunan karya tulis ini. Ucapan terima kasih juga disampaikan kepada seluruh dosen yang telah memberikan landasan ilmu.

#### REFERENSI

- [1] Dinas Lingkungan Hidup Kota Bandung. (2024). Laporan Tahunan Kualitas Udara Kota Bandung Tahun 2024. Bandung: DLH Kota Bandung.
- [2] Badan Pusat Statistik Kota Bandung. (2024). Kota Bandung dalam Angka 2024. Bandung: BPS Kota Bandung.
- [3] Pemerintah Kota Bandung. (2024). Portal Data Terbuka Bandung. Diakses dari <https://data.bandung.go.id> pada 10 Maret 2025.
- [4] Umri, S. S. A., & Umri, S. S. A. (2022). Analisis dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di DKI JAKARTA. *JIKO (Jurnal Informatika Dan Komputer)*, 5(2), 89-96. <https://doi.org/10.33387/jiko.v5i2.2871>
- [5] Ayu, G., Lestari, N., Agus, K., & Aryanto, A. (2023). Peningkatan Akurasi Klasifikasi Kualitas Udara melalui Oversampling dengan Metode Support Vector Machine dan Random Forest. *Jurnal Sistem Dan Informatika (JSI)*, 18(1), 45-53. <https://doi.org/10.30864/jsi.v18i1.596>
- [6] Putri, L. A., & Suwanda. (2023). Implementasi Metode Artificial Neural Network (ANN) Algoritma Backpropagation untuk Klasifikasi Kualitas Udara di Provinsi DKI Jakarta Tahun 2021. *Bandung Conference Series: Statistics*, 3(2), 112-120. <https://doi.org/10.29313/bcss.v3i2.7826>
- [7] Jayadi, B. V., Lauro, M. D., Rusdi, Z., & Handhayani, T. (2024). Air Quality Index Classification for Imbalanced Data using Machine Learning Approach. *Sistemasi: Jurnal Sistem Informasi*, 13(3), 951-958.
- [8] Luthfi, A. M., & Fauzi, F. (2024). Perbandingan Klasifikasi Random Forest, Support Vector Machines, dan LGBM Pada Klasifikasi Kualitas Udara di Jakarta. *Justindo (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 9(2), 78-86. <https://doi.org/10.32528/justindo.v9i2.1912>
- [9] Kementerian Lingkungan Hidup dan Kehutanan. (2023). Peraturan Menteri Lingkungan Hidup dan Kehutanan tentang Indeks Standar Pencemaran Udara. Jakarta: KLHK.
- [10] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- [11] Sajiwo, A. F. B., Rahmat, B., & Junaidi, A. (2024). Klasifikasi Indeks Standar Pencemaran Udara (ISPU) Menggunakan Algoritma Xgboost dengan Teknik Imbalanced Data (SMOTE). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3), 234-242. <https://doi.org/10.23960/jitet.v12i3.4699>
- [12] Rahman, A., & Hidayat, R. (2024). Analisis Perbandingan Kinerja Algoritma Machine Learning untuk Klasifikasi Kualitas Udara di Kota Besar Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 11(4), 721-730. <https://doi.org/10.25126/jtiik.2024114789>