

ANALISIS PENGARUH FAKTOR METEOROLOGI TERHADAP KONSENTRASI PM2.5 GLOBAL DAN IMPLIKASINYA TERHADAP KESEHATAN MASYARAKAT MENGGUNAKAN RANDOM FOREST DAN XGBOOST

Fathoni, Jovandra Anata Malayadza, Fitri Amanda,
Audya Anggraini Hasim, Winda Kurnia Sari, Ken Ditha Tania
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang - Prabumulih No.KM. 32, Indralaya Indah,
Kec. Indralaya, Kabupaten Ogan Ilir, Sumatera Selatan, Indonesia
fathoni@unsri.ac.id

ABSTRAK

PM2.5 merupakan partikel udara berdiameter $\leq 2,5 \mu\text{m}$ yang bersumber dari emisi kendaraan, industri, dan pembakaran biomassa, dan telah menjadi ancaman kesehatan global yang berkontribusi terhadap jutaan kematian akibat penyakit kardiovaskular dan pernapasan. Hubungan antara faktor meteorologi dan konsentrasi PM2.5 bersifat non-linear dan kompleks sehingga tidak dapat dimodelkan secara memadai menggunakan pendekatan statistik konvensional, sementara kajian integratif skala global yang menghubungkan meteorologi, PM2.5, dan implikasi kesehatan masih terbatas. Penelitian ini bertujuan menganalisis pengaruh faktor meteorologi terhadap konsentrasi PM2.5 secara global serta mengevaluasi implikasinya terhadap kesehatan masyarakat berdasarkan standar WHO AQG 2021. Analisis dilakukan menggunakan Random Forest dan XGBoost terhadap Global Air Quality Data yang mencakup 10.000 observasi dari 20 kota di 19 negara dengan variabel input suhu, kelembapan, dan kecepatan angin. Meskipun kemampuan prediksi kuantitatif terbatas akibat kompleksitas data global, Random Forest menunjukkan performa lebih baik dibanding XGBoost dengan MAE 36,31 dan RMSE 42,45. Kecepatan angin, kelembapan, dan suhu teridentifikasi sebagai faktor paling berpengaruh, dengan 79% data berada pada kategori tidak sehat hingga berbahaya berdasarkan WHO AQG 2021.

Kata kunci : Faktor Meteorologi; Kesehatan Masyarakat; Kualitas Udara; PM2.5; Random Forest; XGBoost

1. PENDAHULUAN

Polusi udara merupakan salah satu tantangan lingkungan dan kesehatan terbesar yang dihadapi dunia saat ini. Pencemaran udara terjadi akibat masuknya berbagai zat berbahaya ke dalam atmosfer yang berasal dari aktivitas manusia maupun proses alam [1].

Sumber antropogenik utama meliputi emisi kendaraan bermotor, kegiatan industri, pembangkit listrik berbahan bakar fosil, serta pembakaran biomassa, sedangkan sumber alami antara lain letusan gunung berapi, kebakaran hutan, dan debu mineral. Seiring meningkatnya industrialisasi dan urbanisasi global, kualitas udara di berbagai wilayah terus mengalami penurunan. Kondisi ini menjadikan polusi udara sebagai salah satu risiko lingkungan utama yang berkontribusi terhadap meningkatnya penyakit pernapasan, kardiovaskular, dan berbagai gangguan kesehatan lainnya [1].

PM2.5 merupakan partikel halus berdiameter $\leq 2,5 \mu\text{m}$ yang termasuk polutan udara paling berbahaya bagi kesehatan manusia. Ukurannya yang sangat kecil memungkinkan partikel ini menembus saluran pernapasan hingga mencapai alveoli paru-paru dan bahkan masuk ke aliran darah. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa polusi udara menyebabkan sekitar 7 juta kematian setiap tahun dengan paparan PM2.5 sebagai salah satu faktor risiko utama [2].

Selain itu, analisis Global Burden of Disease (GBD) 2021 menunjukkan bahwa paparan PM2.5

berkontribusi terhadap lebih dari 231 juta Disability-Adjusted Life Years (DALYs) secara global pada tahun 2021 [3].

Besarnya dampak tersebut menunjukkan pentingnya memahami faktor-faktor yang mempengaruhi variasi konsentrasi PM2.5 di atmosfer.

Beberapa penelitian sebelumnya menunjukkan bahwa variabel meteorologi seperti suhu, kelembapan, dan kecepatan angin memiliki pengaruh terhadap variasi konsentrasi PM2.5 di atmosfer [4].

Selain itu, paparan PM2.5 yang tinggi diketahui berhubungan dengan meningkatnya risiko gangguan kesehatan pada masyarakat perkotaan [5].

Hubungan antara faktor cuaca dan konsentrasi PM2.5 juga terbukti berkorelasi dengan peningkatan kasus ISPA [6].

Di sisi lain, pola variasi harian PM2.5 serta keterkaitannya dengan variabel meteorologi menunjukkan dinamika temporal yang kompleks dan tidak selalu linier [7].

Temuan-temuan tersebut mengindikasikan bahwa hubungan antara faktor meteorologi, konsentrasi PM2.5, dan dampaknya terhadap kesehatan bersifat saling terkait dan kompleks.

Meskipun berbagai penelitian tersebut telah mengkaji hubungan antara PM2.5, faktor meteorologi, serta dampaknya terhadap kesehatan, sebagian besar penelitian masih berfokus pada analisis hubungan secara parsial, terbatas pada lokasi tertentu, serta menggunakan pendekatan statistik konvensional. Selain itu, hubungan non-linear dan interaksi

kompleks antar variabel meteorologi dan konsentrasi PM2.5 belum sepenuhnya dapat dijelaskan secara optimal. Integrasi antara analisis faktor meteorologi, dinamika konsentrasi PM2.5, dan implikasinya terhadap kesehatan masyarakat dalam satu kerangka analisis yang komprehensif juga masih relatif terbatas. Oleh karena itu, diperlukan pendekatan yang lebih integratif dan adaptif untuk memahami pola fluktuasi PM2.5 secara lebih menyeluruh.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan penerapan metode machine learning berbasis ensemble, yaitu Random Forest dan Extreme Gradient Boosting (XGBoost), untuk menganalisis pengaruh faktor meteorologi terhadap konsentrasi PM2.5 pada skala global. Pendekatan ini dipilih karena kemampuannya dalam menangkap hubungan non-linear serta mengakomodasi interaksi variabel yang kompleks, sehingga dinilai lebih adaptif dibandingkan pendekatan statistik konvensional [1].

Penelitian ini bertujuan untuk menganalisis pengaruh faktor meteorologi terhadap konsentrasi PM2.5 secara global menggunakan metode Random Forest dan XGBoost serta membandingkan performa kedua model dalam memprediksi konsentrasi PM2.5. Selain itu, penelitian ini juga bertujuan mengevaluasi implikasi hasil prediksi terhadap kesehatan masyarakat berdasarkan standar WHO Air Quality Guidelines (AQG) 2021. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan penerapan machine learning di bidang lingkungan dan kesehatan masyarakat serta menjadi dasar pertimbangan dalam pengendalian polusi udara berbasis data.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian terdahulu telah mengkaji pengaruh faktor meteorologi terhadap PM2.5 dan penerapan machine learning dalam analisisnya. Meneliti sensitivitas PM2.5 terhadap variabel meteorologi di kota-kota tercemar China menggunakan simulasi numerik [8].

Membandingkan delapan algoritma machine learning untuk prediksi PM2.5 berbasis data meteorologi di enam aglomerasi urban China [9].

Mengklasifikasikan konsentrasi PM2.5 menggunakan seleksi fitur dan beberapa model machine learning di Nairobi, Kenya [1].

Melakukan tinjauan sistematis terhadap 79 studi estimasi PM2.5 berbasis machine learning dan data satelit dari berbagai wilayah global [10].

Menganalisis variasi temporal PM2.5 dan kaitannya dengan faktor meteorologi di berbagai kota Indonesia [7].

Meskipun studi-studi tersebut telah memberikan kontribusi penting, kajian yang secara khusus mengintegrasikan Random Forest dan XGBoost untuk menganalisis pengaruh faktor meteorologi terhadap PM2.5 pada skala global beserta implikasinya

terhadap kesehatan masyarakat masih terbatas, sehingga penelitian ini hadir untuk mengisi celah tersebut.

2.2. Partikel Halus PM2.5 dan Distribusi Global

PM2.5 adalah partikel udara berdiameter $\leq 2,5$ μm yang bersumber dari emisi kendaraan, industri, dan pembakaran biomassa. Secara global, kajian kohort internasional membuktikan bahwa PM2.5 secara luas melampaui batas aman WHO sebesar $5 \mu\text{g}/\text{m}^3$ dan berdampak pada kesehatan populasi di berbagai wilayah dunia dengan tingkat paparan yang berbeda-beda [11].

Studi pada kota-kota paling tercemar di dunia mengonfirmasi bahwa konsentrasi PM2.5 dipengaruhi kondisi lokal seperti topografi, kepadatan emisi, serta dinamika atmosfer yang berbeda-beda antar wilayah dan musim, sehingga pendekatan pemodelan yang mempertimbangkan variasi ini menjadi sangat penting [8].

2.3. Pengaruh Faktor Meteorologi terhadap Konsentrasi PM2.5

Faktor meteorologi merupakan penentu utama variasi konsentrasi PM2.5 di seluruh dunia. Simulasi numerik pada kota-kota paling tercemar secara global membuktikan bahwa kecepatan angin adalah variabel paling dominan yang menekan PM2.5 melalui proses dispersi, diikuti oleh PBLH dan curah hujan, sementara suhu dan kelembapan menunjukkan pengaruh yang lebih kompleks dan tidak selalu linear [8].

Pada skala multi-kota, suhu terbukti menjadi faktor meteorologi paling signifikan secara musiman dengan konsentrasi PM2.5 tertinggi pada musim dingin akibat kondisi lapisan batas yang lebih stabil [12].

Di tingkat lokal, pembentukan inversi suhu dan lapisan batas stabil pada malam hari menyebabkan akumulasi polutan secara signifikan, sedangkan pencampuran vertikal pada siang hari menurunkan konsentrasi PM2.5 [13].

Di Indonesia, variasi PM2.5 sangat dipengaruhi pergantian musim dan PBLH, dengan konsentrasi ekstrem yang dipicu kebakaran gambut dan diperparah kondisi meteorologi tertentu [7].

2.4. PM2.5 terhadap Kesehatan Masyarakat

Paparan PM2.5 berdampak serius terhadap kesehatan masyarakat secara global. Meta-analisis terhadap jutaan partisipan membuktikan bahwa paparan jangka panjang PM2.5 meningkatkan risiko mortalitas kardiovaskular secara signifikan, dengan efek yang lebih kuat pada kelompok rentan seperti lansia [14]. Kajian kohort internasional mengonfirmasi bahwa peningkatan konsentrasi PM2.5 konsisten meningkatkan risiko kematian akibat penyakit kardiovaskular dan respirasi di berbagai wilayah dunia [11].

Implikasi kebijakan dari temuan ini sangat nyata: ribuan kematian per tahun dapat dicegah apabila konsentrasi PM_{2.5} memenuhi standar WHO, menegaskan urgensi sistem pemantauan dan analisis kualitas udara berbasis data yang akurat [15].

Lebih jauh, PM_{2.5} secara sinergis dengan perubahan iklim juga memperburuk penyakit pernapasan seperti asma dan PPOK, di mana faktor meteorologi seperti suhu dan gelombang panas turut memperkuat dampak polutan tersebut [15].

2.5. Random Forest dan XGBoost dalam Analisis PM_{2.5}

Random Forest dan *XGBoost* secara konsisten menunjukkan performa tinggi dalam analisis *driver* PM_{2.5} berbasis faktor meteorologi. *Random Forest* mencapai R^2 hingga 0,91 dalam estimasi PM_{2.5} berbasis data satelit [10].

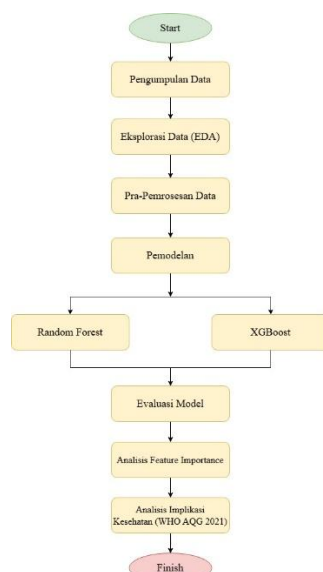
XGBoost mencapai akurasi klasifikasi tertinggi dibanding model lain dalam kategorisasi konsentrasi PM_{2.5}, dengan kelembapan sebagai prediktor meteorologi paling dominan [1].

Studi komparatif delapan model *machine learning* mengonfirmasi bahwa keduanya memberikan kinerja yang stabil dan interpretatif di berbagai aglomerasi urban [12].

Penggunaan metode *Explainable AI* seperti SHAP pada kedua algoritma memungkinkan identifikasi kontribusi relatif setiap variabel meteorologi secara transparan, sehingga hasil analisis dapat langsung diterjemahkan menjadi rekomendasi kebijakan pengendalian kualitas udara dan mitigasi risiko kesehatan masyarakat [12][1].

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini mencakup serangkaian tahapan sistematis yang disusun untuk menjawab rumusan masalah penelitian. Alur penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Flowchart alur penelitian

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif analitik komparatif menggunakan data sekunder berskala global yang bersifat retrospektif. Pendekatan yang diadopsi adalah pemodelan *machine learning supervised* menggunakan dua algoritma *ensemble* berbasis pohon keputusan, yaitu *Random Forest* dan *XGBoost*. Kedua algoritma ini terbukti mampu mempelajari pola *non-linear* yang kompleks antar variabel atmosfer dalam menghasilkan prediksi PM_{2.5} dengan akurasi tinggi [9].

3.2. Sumber dan Karakteristik Data

Data bersumber dari *dataset* Global Air Quality Data yang diunduh dari platform Kaggle (<https://www.kaggle.com>), terdiri dari 10.000 baris data dan 12 variabel yang mencakup 20 kota di 19 negara. Variabel meliputi informasi lokasi (*City*, *Country*), waktu (*Date*), konsentrasi polutan (PM_{2.5}, PM₁₀, NO₂, SO₂, CO, O₃), dan parameter meteorologi (*Temperature*, *Humidity*, *Wind Speed*). Variabel target adalah PM_{2.5}, sedangkan variabel lainnya digunakan sebagai fitur *input*.

3.3. Tahap Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan secara sistematis menggunakan Python di Google Colaboratory, meliputi tahapan berikut:

- Pembersihan data: penghapusan duplikat dan imputasi *missing values* menggunakan median per kota.
- Penanganan *outlier*: *winsorizing* pada batas 1%–99% untuk seluruh variabel numerik.
- Pemisahan fitur dan target: PM_{2.5} sebagai target, variabel lainnya sebagai prediktor.
- Pembagian data: rasio 80:20 secara acak menggunakan fungsi *train_test_split* dari *scikit-learn*.

3.4. Pemodelan Machine Learning

Pemodelan dilakukan menggunakan dua algoritma regresi berbasis pohon keputusan, yaitu *Random Forest Regressor* dan *XGBoost Regressor*, yang diimplementasikan menggunakan *library scikit-learn* dan *XGBoost* dalam lingkungan Python di Google Colaboratory. Kedua algoritma dipilih karena kemampuannya menangkap hubungan *non-linear* yang kompleks antarvariabel dan menghasilkan analisis *feature importance* yang dapat mengukur kontribusi relatif setiap variabel terhadap konsentrasi PM_{2.5} [12] [1].

Model *Random Forest* dikonfigurasi dengan pohon yang dibiarkan tumbuh penuh (*max_depth=None*) dengan parameter *n_estimators=200*. Model *XGBoost* dikonfigurasi dengan parameter *learning_rate=0,05*, *n_estimators=300*, *max_depth=8*, *subsample=0,8* dan *colsample_bytree=0,8*. Kedua model dilatih menggunakan fungsi *fit()* pada data latih, kemudian digunakan untuk memprediksi PM_{2.5} pada data uji.

3.5. Evaluasi Model

Kinerja masing-masing model dievaluasi menggunakan tiga metrik regresi standar: *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). MAE mengukur rata-rata kesalahan absolut, RMSE memberikan penalti lebih besar terhadap kesalahan yang besar, sedangkan R^2 menunjukkan proporsi variansi target yang dapat dijelaskan model. Kedua model dibandingkan langsung berdasarkan ketiga metrik tersebut.

3.6. Visualisasi dan Interpretasi Hasil

Visualisasi hasil dilakukan menggunakan Matplotlib dan Seaborn, meliputi: (1) *heatmap* korelasi antar variabel meteorologi dan PM2.5; (2) *scatter plot* hubungan masing-masing faktor meteorologi terhadap PM2.5; (3) grafik *feature importance* dari kedua model sebagai luaran analisis utama; dan (4) grafik perbandingan rata-rata PM2.5 per negara terhadap ambang batas WHO AQG 2021.

3.7. Analisis Implikasi Kesehatan Masyarakat

Analisis implikasi kesehatan dilakukan secara deskriptif-komparatif dengan membandingkan

konsentrasi PM2.5 aktual terhadap standar WHO Air Quality Guidelines (AQG) 2021 (batas tahunan 5 $\mu\text{g}/\text{m}^3$, batas harian 15 $\mu\text{g}/\text{m}^3$). Data dikategorikan ke dalam lima kelas risiko: Aman ($\leq 5 \mu\text{g}/\text{m}^3$), Moderat ($5\text{--}15 \mu\text{g}/\text{m}^3$), Tidak Sehat Sensitif ($15\text{--}35 \mu\text{g}/\text{m}^3$), Tidak Sehat ($35\text{--}75 \mu\text{g}/\text{m}^3$), dan Berbahaya ($>75 \mu\text{g}/\text{m}^3$). Hasil dikaitkan dengan faktor meteorologi dominan yang teridentifikasi dari analisis *feature importance*.

4. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan analisis pengaruh faktor meteorologi terhadap konsentrasi PM2.5 global menggunakan algoritma *Random Forest* dan *XGBoost*. Kedua model diterapkan untuk mengidentifikasi faktor meteorologi yang paling berkontribusi terhadap konsentrasi PM2.5 serta implikasinya terhadap kesehatan masyarakat berdasarkan standar WHO AQG 2021.

4.1. Deskripsi Dataset

Eksplorasi awal terhadap *dataset* menunjukkan variasi konsentrasi PM2.5 yang cukup tinggi antar wilayah. Statistik deskriptif seluruh variabel numerik disajikan pada Tabel 1 berikut.

Tabel 1. Statistik deskriptif variabel *dataset*

Variabel	Count	Mean	Std	Min	Max	Median
PM2.5	10.000	77,45	41,93	5,02	149,98	77,73
PM10	10.000	104,44	55,06	10,00	200,00	103,69
NO2	10.000	52,20	27,32	5,01	100,00	52,10
SO2	10.000	25,34	14,09	1,00	49,99	25,35
CO	10.000	5,05	2,85	0,10	10,00	5,09
O3	10.000	106,03	55,08	10,04	200,00	106,06
Tempe- rature	10.000	14,90	14,44	-10,00	40,00	14,76
Humidi- ty	10.000	55,08	25,98	10,01	99,99	55,08
Wind Speed	10.000	10,23	5,63	0,50	20,00	10,26

Berdasarkan Tabel 1, konsentrasi PM2.5 memiliki rata-rata 77,45 $\mu\text{g}/\text{m}^3$ (std: 41,93; min: 5,02; maks: 149,98 $\mu\text{g}/\text{m}^3$). Nilai ini jauh melampaui ambang batas WHO AQG 2021 baik tahunan (5 $\mu\text{g}/\text{m}^3$) maupun harian (15 $\mu\text{g}/\text{m}^3$), mengindikasikan bahwa seluruh wilayah dalam *dataset* berada pada kondisi kualitas udara yang berisiko terhadap kesehatan masyarakat.

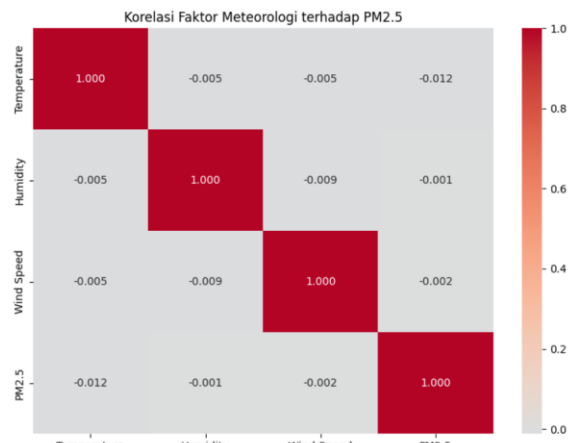
4.2. Hasil Pra-Pemrosesan Data

Hasil pemeriksaan awal menunjukkan bahwa *dataset* tidak mengandung duplikat maupun *missing values* pada seluruh variabel numerik, sehingga tidak diperlukan imputasi tambahan. Proses *winsorizing* (batas 1%–99%) hanya menghasilkan perubahan kecil pada nilai ekstrem PM2.5, dari 149,98 $\mu\text{g}/\text{m}^3$ menjadi 148,66 $\mu\text{g}/\text{m}^3$, menunjukkan distribusi data yang sudah cukup stabil. Pembagian data dengan rasio 80:20 menghasilkan 8.000 sampel data latih dan 2.000 sampel data uji yang selanjutnya digunakan dalam proses pemodelan.

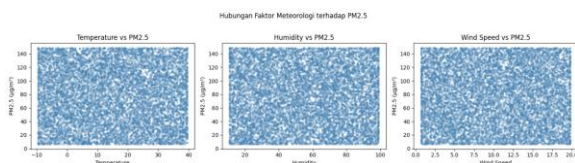
4.3. Hasil Analisis Korelasi Faktor Meteorologi terhadap PM2.5

Analisis korelasi Pearson menunjukkan bahwa ketiga faktor meteorologi memiliki korelasi yang sangat lemah terhadap PM2.5: suhu ($-0,012$), kecepatan angin ($-0,002$), dan kelembapan ($-0,001$). Meskipun nilainya kecil, arah negatif ketiga variabel ini sesuai secara teoritis, peningkatan kecepatan angin dan kelembapan berkaitan dengan penurunan PM2.5 melalui mekanisme dispersi partikel dan *wet deposition* [4][7].

Lemahnya korelasi linear ini mengonfirmasi bahwa hubungan antara faktor meteorologi dan PM2.5 bersifat *non-linear* dan tidak dapat ditangkap secara memadai oleh analisis korelasi konvensional. Hal ini sejalan dengan *scatter plot* yang memperlihatkan sebaran data tanpa pola tren yang jelas, dan memperkuat relevansi penggunaan algoritma berbasis *ensemble* dalam penelitian ini.



Gambar 2. Heatmap korelasi faktor meteorologi terhadap PM2.5



Gambar 3. Scatter plot hubungan faktor meteorologi (suhu, kelembapan, kecepatan angin) terhadap PM2.5

4.4. Hasil Pemodelan dan Evaluasi Model

Hasil evaluasi performa kedua model disajikan pada Tabel 2.

Tabel 2. Perbandingan performa model *random forest* dan *XGBoost*

Model	MAE	RMSE	R ²
<i>Random Forest</i>	36,3123	42,4481	-0,0293
<i>XGBoost</i>	37,0670	43,4556	-0,0788

Random Forest menunjukkan performa lebih baik dibandingkan *XGBoost* pada seluruh metrik, dengan MAE 36,31 $\mu\text{g}/\text{m}^3$, RMSE 42,45 $\mu\text{g}/\text{m}^3$, dan R² -0,0293. Nilai R² yang negatif pada kedua model mengindikasikan bahwa model belum mampu menjelaskan variansi PM2.5 secara memadai melalui prediksi kuantitatif. Kondisi ini konsisten dengan lemahnya korelasi linear antara variabel, dan mengindikasikan bahwa variasi PM2.5 dalam *dataset* global ini dipengaruhi oleh interaksi kompleks yang tidak sepenuhnya tertangkap oleh fitur yang tersedia, seperti faktor emisi lokal, topografi, dan dinamika lapisan atmosfer [9,10]. Meski demikian, kemampuan kedua model dalam menghasilkan analisis *feature importance* tetap valid dan bermakna sebagai alat untuk mengidentifikasi pengaruh relatif masing-masing variabel.

4.5. Analisis Feature Importance: Pengaruh Faktor Meteorologi terhadap PM2.5

Analisis *feature importance* merupakan luaran utama penelitian ini dalam menjawab rumusan masalah mengenai pengaruh faktor meteorologi

terhadap PM2.5. Hasil dari kedua model disajikan pada Tabel 3.

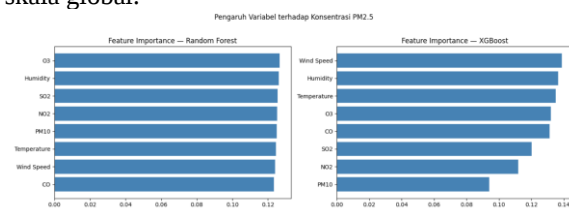
Tabel 3. *Feature importance random forest* dan *XGBoost*

Variabel	RF Importance	XGB Importance	Rata-rata
<i>Wind Speed</i>	0,1241	0,1389	0,1315
<i>Humidity</i>	0,1260	0,1364	0,1312
<i>Temperature</i>	0,1244	0,1352	0,1298
O ₃	0,1265	0,1321	0,1293
CO	0,1234	0,1313	0,1274
SO ₂	0,1255	0,1202	0,1229
NO ₂	0,1251	0,1118	0,1185
PM10	0,1250	0,0941	0,1096

Model *Random Forest* menghasilkan distribusi *feature importance* yang sangat merata (0,123–0,127), mengindikasikan bahwa seluruh variabel berkontribusi secara relatif seimbang tanpa satu variabel yang dominan. Pola ini mencerminkan karakteristik *Random Forest* yang membangun banyak pohon secara paralel dengan pemilihan fitur secara acak, sehingga cenderung mendistribusikan kontribusi secara lebih merata [12].

Sebaliknya, *XGBoost* menghasilkan distribusi yang lebih variatif dan informatif. Ketiga faktor meteorologi *Wind Speed* (0,1389), *Humidity* (0,1364), dan *Temperature* (0,1352) secara konsisten menempati tiga peringkat teratas, jauh di atas PM10 (0,0941) yang berada di posisi terendah. Pola ini menunjukkan bahwa model *XGBoost* lebih sensitif dalam membedakan kontribusi antar variabel, yang merupakan keunggulan pendekatan *boosting* dalam menangkap sinyal variabel yang lemah namun konsisten [1].

Jika dirata-ratakan dari kedua model, tiga faktor meteorologi yang paling berpengaruh secara berurutan adalah *Wind Speed* (0,1315), *Humidity* (0,1312), dan *Temperature* (0,1298). Secara mekanistik, kecepatan angin yang tinggi mempercepat dispersi dan pengenceran partikel PM2.5 di atmosfer. Kelembapan yang tinggi mendorong proses *wet deposition* dan pertumbuhan higroskopik partikel yang memperberat deposisinya. Sementara itu, suhu memengaruhi laju reaksi fotokimia yang berperan dalam pembentukan partikel sekunder. Secara keseluruhan, ketiga faktor meteorologi ini terbukti memiliki kontribusi paling besar dibandingkan variabel polutan lainnya, mengindikasikan bahwa kondisi atmosfer memainkan peran penting dalam variasi konsentrasi PM2.5 pada skala global.

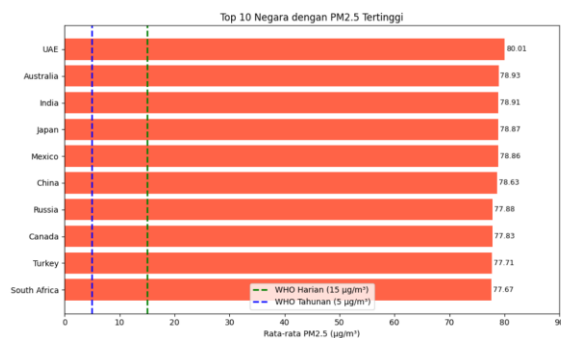


Gambar 4. *Feature importance random forest* dan *XGBoost* terhadap konsentrasi PM2.5

4.6. Implikasi terhadap Kesehatan Masyarakat

Kategorisasi konsentrasi PM2.5 aktual berdasarkan WHO AQG 2021 menunjukkan hasil yang mengkhawatirkan: dari 10.000 data, sebanyak 5.190 (51,9%) masuk kategori Berbahaya ($>75 \mu\text{g}/\text{m}^3$), 2.733 (27,3%) Tidak Sehat ($35\text{--}75 \mu\text{g}/\text{m}^3$), 1.368 (13,7%) Tidak Sehat Sensitif ($15\text{--}35 \mu\text{g}/\text{m}^3$), dan 709 (7,1%) Moderat ($5\text{--}15 \mu\text{g}/\text{m}^3$). Tidak ada satu pun data yang berada pada kategori Aman ($\leq 5 \mu\text{g}/\text{m}^3$), menunjukkan bahwa seluruh wilayah dalam *dataset* terpapar PM2.5 di atas batas aman WHO.

Secara spasial, Uni Emirat Arab (80,01 $\mu\text{g}/\text{m}^3$), Australia (78,93), India (78,91), Jepang (78,87), dan Meksiko (78,86 $\mu\text{g}/\text{m}^3$) mencatat rata-rata PM2.5 tertinggi. Menariknya, keragaman geografis negara-negara ini mulai dari kawasan gurun, subtropis, hingga Asia Selatan mengindikasikan bahwa tingginya PM2.5 tidak semata-mata dipengaruhi oleh iklim, melainkan juga oleh intensitas aktivitas industri dan transportasi di masing-masing wilayah.



Gambar 5. Rata-rata konsentrasi PM2.5 per negara dibandingkan ambang batas WHO AQG 2021

Dikaitkan dengan hasil *feature importance*, rendahnya kecepatan angin dan kelembapan di wilayah-wilayah tersebut berpotensi menghambat proses dispersi dan deposisi alami partikel, sehingga memperparah akumulasi PM2.5. Paparan kronis pada level berbahaya ini berisiko tinggi terhadap sistem pernapasan dan kardiovaskular, khususnya bagi kelompok rentan seperti lansia, anak-anak, dan penderita penyakit kronis [2] [14]. Oleh karena itu, pemantauan kondisi meteorologi secara *real-time* dan integrasi data cuaca dalam sistem peringatan dini kualitas udara menjadi langkah strategis yang perlu dipertimbangkan dalam kebijakan kesehatan lingkungan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis menggunakan algoritma Random Forest dan XGBoost terhadap data kualitas udara global, dapat disimpulkan bahwa faktor meteorologi memiliki pengaruh terhadap variasi konsentrasi PM2.5, meskipun hubungan yang terbentuk bersifat kompleks dan *non-linear*. Hasil evaluasi model menunjukkan bahwa *Random Forest* memiliki performa prediksi yang sedikit lebih baik dibandingkan XGBoost berdasarkan nilai MAE dan

RMSE. Analisis *feature importance* dari kedua model mengindikasikan bahwa kecepatan angin, kelembapan, dan suhu merupakan faktor meteorologi yang paling berpengaruh terhadap konsentrasi PM2.5. Selain itu, hasil kategorisasi berdasarkan standar WHO AQG 2021 menunjukkan bahwa sebagian besar data berada pada kategori tidak sehat hingga berbahaya, yang menandakan tingginya potensi risiko kesehatan masyarakat akibat paparan polusi udara partikulat halus. Oleh karena itu, penelitian ini menegaskan pentingnya integrasi data meteorologi dalam analisis kualitas udara untuk mendukung sistem pemantauan dan pengendalian polusi udara berbasis data.

Untuk penelitian selanjutnya, disarankan menggunakan *dataset* dengan cakupan variabel yang lebih luas, seperti data emisi industri, kepadatan transportasi, dan kondisi topografi, serta mengembangkan pendekatan model yang lebih kompleks atau berbasis *deep learning* agar mampu meningkatkan akurasi prediksi konsentrasi PM2.5 secara lebih optimal.

DAFTAR PUSTAKA

- [1] C. M. Matara, S. O. Nyambane, A. O. Yusuf, E. A. Ochungo, and A. Khattak, "Classification of Particulate Matter (PM2.5) Concentrations Using Feature Selection and Machine Learning Strategies," *LOGI - Scientific Journal on Transport and Logistics*, vol. 15, no. 1, pp. 85–96, Jan. 2024, doi: 10.2478/logi-2024-0008.
- [2] Nafas, DBS Foundation, and U. of I. (FKM U. Faculty of Public Health, "The Impact of PM2.5 on Pneumonia in Toddlers in Jakarta," 2025.
- [3] T. Fang *et al.*, "Temporal trends of particulate matter pollution and its health burden, 1990–2021, with projections to 2036: a systematic analysis for the global burden of disease study 2021," *Front. Public Health*, vol. 13, 2025, doi: 10.3389/fpubh.2025.1579716.
- [4] I. R. Qonita, I. P. Santikayasa, and A. Turyanti, "Analisis Pengaruh Faktor Meteorologi terhadap Fluktuasi PM2.5 di Kota Pontianak," *POSITRON*, vol. 15, no. 1, pp. 24–33, May 2025, doi: 10.26418/positron.v15i1.91566.
- [5] E. Lolita and E. H. Pakpahan, "HUBUNGAN PAPARAN POLUSI UDARA PM2,5 DENGAN RISIKO GANGGUAN KESEHATAN MASYARAKAT DI WILAYAH PERKOTAAN," *Journal of Public Health Sciences*, vol. 14, no. 2, pp. 254–262, 2025, [Online]. Available: <https://jurnal.ikta.ac.id/index.php/kesmas>
- [6] C. Rahadiyanto, "ANALISIS PENGARUH CUACA DAN PM2.5 TERHADAP PREVALENSI ISPA DI KOTA SEMARANG DENGAN PENDEKATAN STATISTIK DAN MACHINE LEARNING," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11,

- no. 1, pp. 934–948, Jan. 2026, doi: 10.36341/rabit.v11i1.7190.
- [7] N. Cholianawati *et al.*, “Diurnal and Daily Variations of PM_{2.5} and its Multiple-Wavelet Coherence with Meteorological Variables in Indonesia,” *Aerosol Air Qual. Res.*, vol. 24, no. 3, Mar. 2024, doi: 10.4209/aaqr.230158.
- [8] J. Wang, J. Han, T. Li, T. Wu, and C. Fang, “Impact analysis of meteorological variables on PM_{2.5} pollution in the most polluted cities in China,” *Heliyon*, vol. 9, no. 7, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17609.
- [9] A. Makhdoomi, M. Sarkhosh, and S. Ziaei, “PM_{2.5} concentration prediction using machine learning algorithms: an approach to virtual monitoring stations,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-92019-3.
- [10] M. Unik, I. S. Sitanggang, L. Syaufina, I. Nengah, and S. Jaya, “PM 2.5 Estimation using Machine Learning Models and Satellite Data: A Literature Review.” [Online]. Available: www.ijacsa.thesai.org
- [11] S. Chen *et al.*, “Global associations between long-term exposure to PM_{2.5} constituents and health: A systematic review and meta-analysis of cohort studies,” Aug. 05, 2024, *Elsevier B.V.* doi: 10.1016/j.jhazmat.2024.134715.
- [12] M. Duan, Y. Sun, B. Zhang, C. Chen, T. Tan, and Y. Zhu, “PM_{2.5} Concentration Prediction in Six Major Chinese Urban Agglomerations: A Comparative Study of Various Machine Learning Methods Based on Meteorological Data,” *Atmosphere (Basel)*, vol. 14, no. 5, May 2023, doi: 10.3390/atmos14050903.
- [13] A. Samad *et al.*, “Meteorological and air quality measurements in a city region with complex terrain: influence of meteorological phenomena on urban climate,” *Meteorologische Zeitschrift*, vol. 32, no. 4, pp. 293–315, 2023, doi: 10.1127/metz/2023/1124.
- [14] C. Krittanawong *et al.*, “PM_{2.5} and Cardiovascular Health Risks,” Jun. 01, 2023, *Elsevier Inc.* doi: 10.1016/j.cpcardiol.2023.101670.
- [15] H. M. Tran *et al.*, “The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence,” Nov. 10, 2023, *Elsevier B.V.* doi: 10.1016/j.scitotenv.2023.166340