

Prediksi Konsentrasi PM2.5 Berbasis Ensemble Machine Learning di Kota Pontianak

Prediction of PM2.5 Concentration Using Ensemble Machine Learning in Pontianak City

Irfana Diah Faryuni^{1*}, Joko Sampurno¹

¹Program Studi Geofisika, Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura, Pontianak, Indonesia 78234

Email: *irfana@physics.untan.ac.id; jokosampurno@physics.untan.ac.id

Received: 17th March 2026; Revised: 21th May 2026; Accepted: 22nd July 2026

Abstrak – Konsentrasi partikulat halus (PM2.5) merupakan indikator utama kualitas udara yang berdampak signifikan terhadap kesehatan manusia dan lingkungan. Kota Pontianak sebagai wilayah tropis ekuatorial menghadapi tantangan serius terkait fluktuasi konsentrasi PM2.5 akibat pengaruh meteorologi lokal dan potensi kebakaran lahan di wilayah sekitarnya. Penelitian ini bertujuan untuk mengembangkan model prediksi per jam konsentrasi PM2.5 menggunakan pendekatan ensemble machine learning berbasis kombinasi Random Forest dan Extreme Gradient Boosting (XGBoost). Data yang digunakan mencakup observasi jam-jaman selama satu tahun (Agustus 2024–Agustus 2025) yang terdiri atas konsentrasi PM2.5 serta variabel meteorologi (suhu, kelembapan, dan tekanan udara). Analisis korelasi dan mutual information menunjukkan bahwa fitur temporal, khususnya lagged PM2.5 dan rolling average, memiliki kontribusi paling dominan terhadap nilai PM2.5 saat ini, sedangkan pengaruh faktor meteorologi relatif kecil dan bersifat tidak langsung. Hasil pemodelan menunjukkan bahwa model Hybrid Ensemble (Stacking RF–XGB) memberikan performa terbaik dengan nilai $R^2 = 0,72$, $MAE = 9,71 \mu\text{g}/\text{m}^3$, dan $RMSE = 22,79 \mu\text{g}/\text{m}^3$, lebih unggul dibandingkan dengan model tunggal Random Forest ($R^2 = 0,69$) dan XGBoost ($R^2 = 0,65$). Temuan ini menegaskan bahwa kombinasi model ensemble mampu menangkap pola nonlinier dan hubungan temporal kompleks secara lebih efektif. Pendekatan ini berpotensi diterapkan sebagai dasar pengembangan sistem prediksi dan peringatan dini kualitas udara di kawasan tropis, khususnya wilayah rawan polusi seperti Pontianak.

Kata kunci: PM2.5; ensemble machine learning; random forest; XGBoost; prediksi kualitas udara.

Abstract – Fine particulate matter (PM2.5) is a key indicator of air quality, profoundly impacting human health and the environment. Pontianak City, located in the equatorial tropics of Indonesia, faces recurring air quality challenges driven by local meteorological variability and seasonal biomass burning in the surrounding regions. This study developed an hourly PM2.5 prediction model using an ensemble machine learning approach that integrates Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms. The dataset comprised hourly observations aggregated into hourly values from August 2024 to August 2025, including PM2.5 concentrations and three meteorological predictors: temperature, relative humidity, and atmospheric pressure. Correlation and mutual information analyses revealed that temporal features, notably lagged PM2.5 and 24-hour rolling averages, exerted the strongest influence on current PM2.5 levels, whereas meteorological variables contributed marginally in a nonlinear manner. Model evaluation demonstrated that the Hybrid Ensemble (Stacking RF–XGB) achieved the best performance with $R^2 = 0.72$, $MAE = 9.71 \mu\text{g}/\text{m}^3$, and $RMSE = 22.79 \mu\text{g}/\text{m}^3$, outperforming individual models (RF: $R^2 = 0.69$; XGBoost: $R^2 = 0.65$). The hybrid model effectively captured temporal fluctuations and extreme pollution events, offering improved robustness and generalization. These results highlight the potential of ensemble-based machine learning to enhance short-term air quality forecasting systems in tropical regions, providing valuable support for public health management and early warning strategies in Pontianak and similar urban environments in the future.

Keywords: PM2.5; Ensemble machine learning; random forest; XGBoost; tropical urban air quality.

1. Pendahuluan

Konsentrasi komponen partikel kecil dengan diameter lebih kecil dari 2,5 mikrometer (*Particulate Matter* (PM2.5)) dalam polusi udara telah diakui memiliki keterkaitan yang kuat dengan berbagai dampak buruk terhadap kesehatan manusia [1]. Partikel halus dengan diameter kurang dari 2,5 mikrometer ini mampu menembus saluran pernapasan hingga ke aliran darah, memicu beragam gangguan serius seperti penyakit pernapasan kronis, gangguan kardiovaskular, hingga peningkatan risiko kematian prematur [2, 3]. Di Indonesia, peningkatan polusi udara perkotaan menjadi tantangan yang semakin besar seiring pertumbuhan populasi, industrialisasi, dan urbanisasi yang pesat, yang secara langsung menempatkan jutaan penduduk pada risiko kesehatan yang signifikan [4].

Secara khusus, Kota Pontianak di Kalimantan Barat menghadapi tantangan unik dan berulang dalam menjaga kualitas udara. Setiap tahun, terutama selama musim kemarau, wilayah ini sangat rentan terhadap kabut asap lintas batas yang bersumber dari kebakaran hutan dan lahan gambut (karhutla) di daerah sekitar [5]. Fenomena ini tidak hanya mengganggu aktivitas sosial-ekonomi masyarakat, tetapi juga menyebabkan peningkatan drastis konsentrasi PM2.5 hingga mencapai level berbahaya bagi kesehatan publik.

Pemantauan kualitas udara secara real-time dapat memberikan informasi tentang kondisi terkini. Namun, data historis saja belum dapat menjelaskan dinamika dan determinan utama yang memengaruhi variasi konsentrasi polutan. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengeksplorasi hubungan kompleks antar variabel serta menilai kontribusi relatif faktor-faktor meteorologi terhadap fluktuasi PM2.5. Dalam konteks ini, metode *machine learning* menjadi alternatif yang unggul karena kemampuannya menangani data berskala besar dan nonlinear [6].

Penelitian ini mengimplementasikan dan membandingkan tiga pendekatan pemodelan, yaitu *Random Forest*, *Extreme Gradient Boosting* (XGBoost) [7], dan model gabungan (*ensemble*) keduanya [8]. Sebelum proses pemodelan, dilakukan analisis korelasi untuk mengidentifikasi keterkaitan linier variabel, serta analisis mutual information untuk menilai fitur yang paling informatif terhadap PM2.5 [9]. Kinerja setiap model, termasuk model *ensemble*, dievaluasi secara kuantitatif menggunakan metrik koefisien determinasi (R^2), MAE, dan RMSE [6, 10–12]. Selain itu, dilakukan pula analisis kontribusi fitur melalui *feature importance* guna memahami variabel meteorologi yang paling berpengaruh terhadap dinamika PM2.5 per jam. Penelitian ini bertujuan untuk: (1) menganalisis keterkaitan antara konsentrasi PM2.5 dan faktor-faktor meteorologi melalui pendekatan korelasi dan mutual information; (2) membangun serta membandingkan model *Random Forest*, *XGBoost*, dan kombinasi *ensemble*-nya untuk prediksi jangka pendek PM2.5; serta (3) mengidentifikasi variabel yang paling menentukan konsentrasi PM2.5 di Kota Pontianak. Pendekatan berbasis *machine learning* telah terbukti efektif dalam memodelkan polusi udara, khususnya PM2.5, di berbagai wilayah [3, 6, 12, 13]. Hasil penelitian ini diharapkan dapat memberikan pemahaman ilmiah mengenai faktor-faktor utama yang mengendalikan polusi partikulat di wilayah tropis ekuatorial.

2. Material dan Metode

2.1 Pengumpulan dan pra-pemrosesan data

Penelitian ini menggunakan data deret waktu berbasis sensor dari stasiun pemantauan kualitas udara yang berlokasi di Kota Pontianak, Kalimantan Barat, dengan koordinat alat sensor berada pada Lintang 0,058° LS, dan Bujur 109,3458° BT). Data kualitas udara diperoleh menggunakan sensor IQAir AirVisual Outdoor sebagai sumber data utama. Sensor ini merupakan perangkat pemantauan kualitas udara berbasis *laser scattering* yang mampu mengukur parameter AQI (*Air Quality Index*), PM2.5, Suhu, Tekanan, dan Kelembapan secara *real-time*. Berdasarkan spesifikasi teknis perangkat, sensor PM2.5 memiliki rentang pengukuran 0-1000 $\mu\text{g}/\text{m}^3$ dengan tingkat akurasi $\pm 10 \mu\text{g}/\text{m}^3$ atau 10% sedangkan sensor kelembapan memiliki akurasi $\pm 1\%$ RH. Data dipantau secara daring melalui platform IQAir Visual sehingga memungkinkan akuisisi data deret waktu dengan resolusi temporal tinggi [14]. Periode pengamatan mencakup satu tahun penuh, yaitu dari 31 Agustus 2024 hingga 31 Agustus 2025, dengan resolusi temporal per jam. Parameter utama yang diamati adalah konsentrasi partikulat halus PM2.5 ($\mu\text{g}/\text{m}^3$) sebagai indikator utama kualitas udara. Selain itu, digunakan pula variabel meteorologi pendukung, yaitu suhu

udara ($^{\circ}\text{C}$), kelembapan relatif (*Relative Humidity* (RH) (%)) , dan tekanan udara (hPa), yang diperoleh dari sumber data yang sama untuk menjaga konsistensi temporal.

Tahap pra-pemrosesan data dilakukan secara bertahap untuk memastikan kualitas dan kesesuaian data sebelum proses analisis. Pertama, dilakukan pemeriksaan kelengkapan data untuk mengidentifikasi keberadaan nilai hilang (*missing values*) atau data anomali. Nilai hilang yang jumlahnya kecil diinterpolasi menggunakan metode interpolasi linear, sedangkan deret waktu yang menunjukkan kesalahan pencatatan dihapus agar tidak memengaruhi kestabilan model. Setelah data bersih, semua variabel kemudian dinormalisasi ke dalam rentang $[0, 1]$ menggunakan metode *Min–Max scaling* untuk menghindari dominasi variabel dengan skala besar dan mempercepat konvergensi model.

2.2. Analisis data dan pemodelan

Analisis diawali dengan penerapan statistik deskriptif dan eksplorasi deret waktu untuk menggambarkan karakteristik distribusi serta dinamika temporal konsentrasi PM2.5 setiap jam di Pontianak. Selanjutnya dilakukan analisis hubungan antara PM2.5 dan faktor meteorologi menggunakan korelasi Pearson untuk mengukur hubungan linier. Selain itu, diterapkan pula analisis *mutual information* (MI) guna mengungkap hubungan nonlinear yang mungkin tersembunyi. Hasil analisis divisualisasikan dalam bentuk *heatmap* dan diagram informasi fitur untuk menunjukkan variabel yang paling berpengaruh terhadap fluktuasi PM2.5.

Tahap berikutnya adalah pemodelan prediktif menggunakan algoritma *Random Forest* (RF), *Extreme Gradient Boosting* (XGBoost), dan model hibrida (*ensemble*) yang menggabungkan keduanya. Pemilihan metode *machine learning* ini didasarkan pada karakteristik data PM2.5 yang bersifat nonlinier, memiliki fluktuasi temporal yang tinggi, serta dipengaruhi oleh interaksi kompleks antara variabel meteorologi dan historis polutan. RF digunakan karena kemampuannya menangani hubungan nonlinear, relatif tahan terhadap *noise* dan *overfitting* melalui mekanisme *bootstrap aggregating* (*bagging*), serta efektif dalam mengidentifikasi kontribusi masing-masing melalui analisis *feature importance*. Sementara itu, XGBoost digunakan karena memiliki kemampuan pembelajaran bertahap melalui teknik *gradient boosting* dan regularisasi, sehingga efektif dalam menangkap pola kompleks dan meningkatkan akurasi prediksi pada data deret waktu dengan variabilitas tinggi [15]. Model *ensemble* dikembangkan dengan pendekatan *stacking*, dimana prediksi yang dihasilkan oleh model RF dan XGBoost pada tahap pertama digunakan kembali sebagai masukan pada tahap kedua untuk menghasilkan prediksi akhir yang lebih optimal. Proses penggabungan dilakukan menggunakan pendekatan rata-rata tertimbang (*weighted average*) untuk mencapai keseimbangan antara bias dan varians, serta meningkatkan kestabilan prediksi. Selain itu, digunakan pula fitur waktu berupa *lag* (waktu tertunda) dan rata-rata bergerak (*rolling mean*) dari PM2.5 itu sendiri guna menangkap efek autokorelasi terhadap dinamika konsentrasi PM2.5.

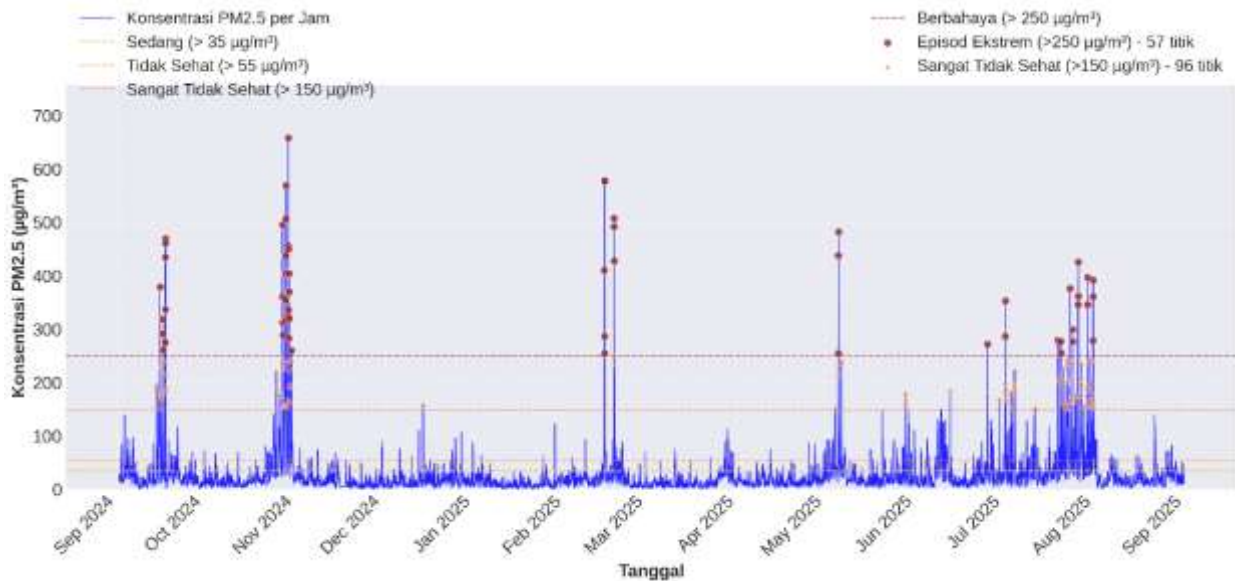
Pembagian dataset dilakukan secara kronologis dengan rasio 90% untuk data latih dan 10% untuk data uji untuk mencegah kebocoran data (*leakage*) dan menjaga kontinuitas temporal. Evaluasi kinerja model dilakukan menggunakan tiga metrik utama, yaitu koefisien determinasi (R^2) untuk mengukur proporsi variasi PM2.5 yang dapat dijelaskan oleh model, *Mean Absolute Error* (MAE) untuk menilai tingkat kesalahan absolut, serta *Root Mean Squared Error* (RMSE) untuk mengevaluasi sensitivitas model terhadap kesalahan besar. Selain pengukuran kuantitatif, dilakukan pula analisis *feature importance* untuk mengidentifikasi variabel yang paling dominan memengaruhi variasi PM2.5.

3. Hasil dan Pembahasan

3.1 Karakteristik PM2.5 dan pola temporal

Berdasarkan hasil pengamatan deret waktu konsentrasi PM2.5 setiap jam selama periode September 2024 hingga September 2025, terlihat adanya fluktuasi yang sangat dinamis dengan sejumlah episode lonjakan ekstrem (Gambar 1). Peningkatan tajam konsentrasi PM2.5 tercatat pada Oktober–November 2024, Maret 2025, serta Juli–Agustus 2025, di mana nilai setiap jam beberapa kali menembus ambang kategori sangat tidak sehat ($>150 \mu\text{g}/\text{m}^3$) hingga berbahaya ($>250 \mu\text{g}/\text{m}^3$). Pola ini mengindikasikan adanya pengaruh kuat kondisi meteorologis dan sumber emisi musiman terhadap variabilitas PM2.5. Periode-periode dengan konsentrasi tinggi tersebut bertepatan dengan fase musim kemarau, ketika curah hujan menurun dan kelembapan relatif rendah, sehingga memperbesar potensi akumulasi partikulat di atmosfer. Selain itu, periode ini juga kerap dikaitkan dengan aktivitas kebakaran hutan dan lahan (karhutla) di Kalimantan yang

menghasilkan kabut asap lintas provinsi dan meningkatkan polusi udara regional. Sebaliknya, pada bulan Desember 2024 hingga Februari 2025, konsentrasi PM2.5 menurun signifikan dan lebih stabil pada rentang kategori sedang hingga tidak sehat ringan, seiring meningkatnya intensitas hujan yang membantu proses *wet deposition* (pencucian atmosfer) serta mempercepat dispersi polutan.



Gambar 1. Tren konsentrasi PM2.5 dengan kategori kualitas udara.

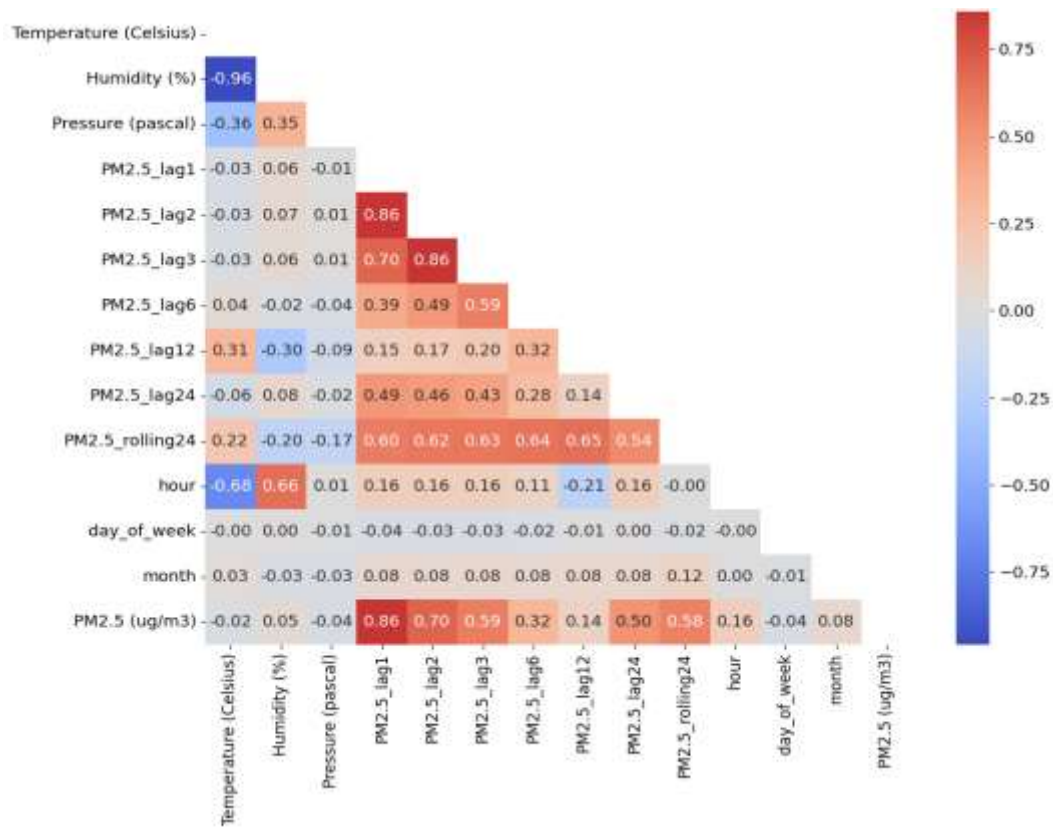
3.2. Hubungan antara PM2.5 dan faktor meteorologi

Hasil analisis korelasi antarvariabel (Gambar 2) menunjukkan bahwa faktor meteorologi seperti suhu udara, kelembapan relatif, dan tekanan permukaan memiliki hubungan yang relatif lemah terhadap konsentrasi PM2.5, dengan koefisien korelasi berada pada kisaran $-0,04$ hingga $0,05$. Di antara ketiganya, kelembapan memiliki korelasi negatif paling kuat terhadap PM2.5 ($r = 0,05$), sedangkan tekanan udara dan suhu menunjukkan hubungan yang sangat lemah dengan nilai PM2.5 setiap jam. Korelasi negatif yang sangat kuat antara suhu dan kelembapan ($r = -0,96$) menandakan adanya keterkaitan fisis yang umum terjadi di daerah tropis, di mana peningkatan suhu biasanya diikuti penurunan kelembapan relatif akibat penguapan intensif. Sebaliknya, variabel lagged PM2.5 memperlihatkan korelasi tinggi dengan nilai PM2.5 saat ini. Korelasi tertinggi tercatat pada lag1 ($r = 0,86$), diikuti lag2 ($r = 0,70$) dan lag3 ($r = 0,59$), kemudian menurun pada jeda waktu yang lebih panjang (lag12 dan lag24). Sementara itu, fitur rata-rata bergulir (*rolling mean*) 24 jam (PM2.5_rolling24) juga memiliki korelasi cukup kuat ($r = 0,58$), menandakan bahwa pola jangka pendek berperan penting dalam menjelaskan variasi PM2.5. Hasil ini menegaskan bahwa informasi temporal atau historis PM2.5 jauh lebih dominan dibandingkan dengan pengaruh langsung faktor meteorologi dalam memprediksi konsentrasi polutan pada setiap jam.

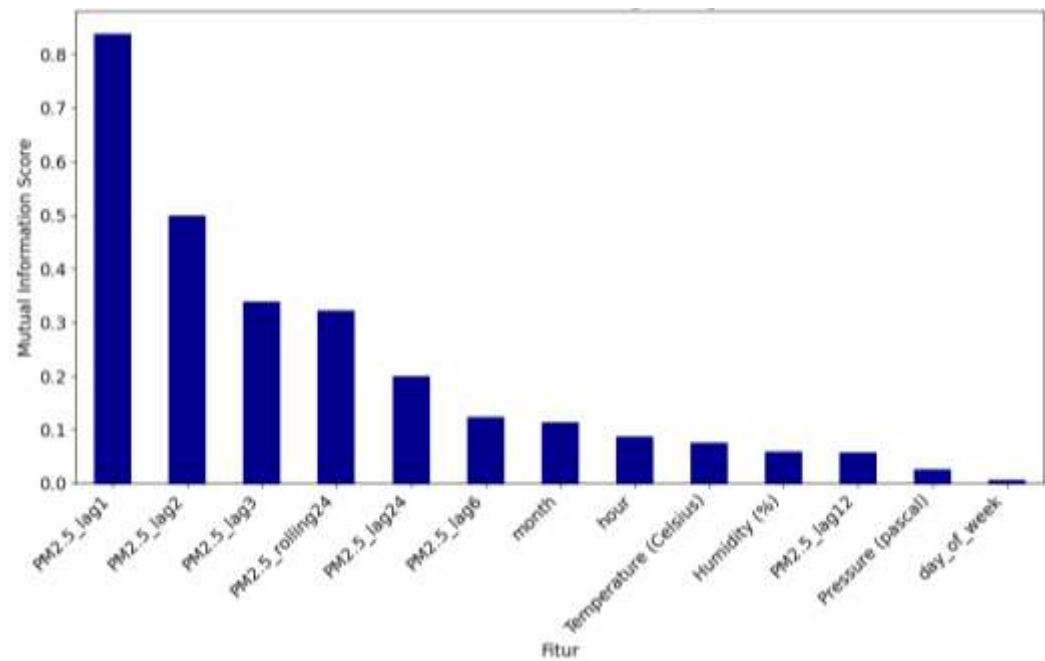
Temuan korelasi tersebut diperkuat melalui hasil analisis *Mutual Information* (MI) (Gambar 3), yang menyoroti besarnya kontribusi setiap fitur terhadap variabilitas PM2.5 tanpa bergantung pada asumsi linearitas hubungan. Fitur PM2.5_lag1 mencatat skor tertinggi ($>0,8$), menunjukkan bahwa kondisi kualitas udara satu jam sebelumnya memiliki pengaruh paling besar terhadap nilai PM2.5 saat ini. Selanjutnya, PM2.5_lag2 menunjukkan skor menengah ($\sim 0,5$), diikuti PM2.5_rolling24 ($\sim 0,32$), yang menandakan bahwa efek akumulatif jangka pendek tetap signifikan. Lag yang lebih jauh seperti lag3 hingga lag6 masih memberikan informasi tambahan meskipun kontribusinya semakin menurun secara bertahap. Sebaliknya, variabel meteorologi yaitu suhu, kelembapan, dan tekanan, memperlihatkan skor informasi yang jauh lebih rendah ($<0,1$), sehingga pengaruhnya terhadap variasi PM2.5 bersifat tidak langsung dan kemungkinan berperan melalui interaksi nonlinear dengan faktor temporal atau emisi.

Konsistensi antara hasil analisis korelasi dan *mutual information* memperkuat kesimpulan bahwa fitur temporal PM2.5 (*lag* dan *rolling mean*) merupakan prediktor utama dalam model prediksi kualitas udara di Pontianak. Dengan demikian, pendekatan berbasis *ensemble machine learning* seperti *Random Forest* dan *XGBoost* dinilai sangat sesuai untuk menangkap hubungan kompleks antara variabel historis dan

lingkungan. Model-model tersebut mampu memanfaatkan pola nonlinear serta interaksi antar fitur yang tidak dapat dijelaskan sepenuhnya oleh pendekatan korelasi konvensional, sehingga diharapkan dapat meningkatkan akurasi estimasi konsentrasi PM2.5 per jam.



Gambar 2. Heatmap korelasi antara fitur dan target.

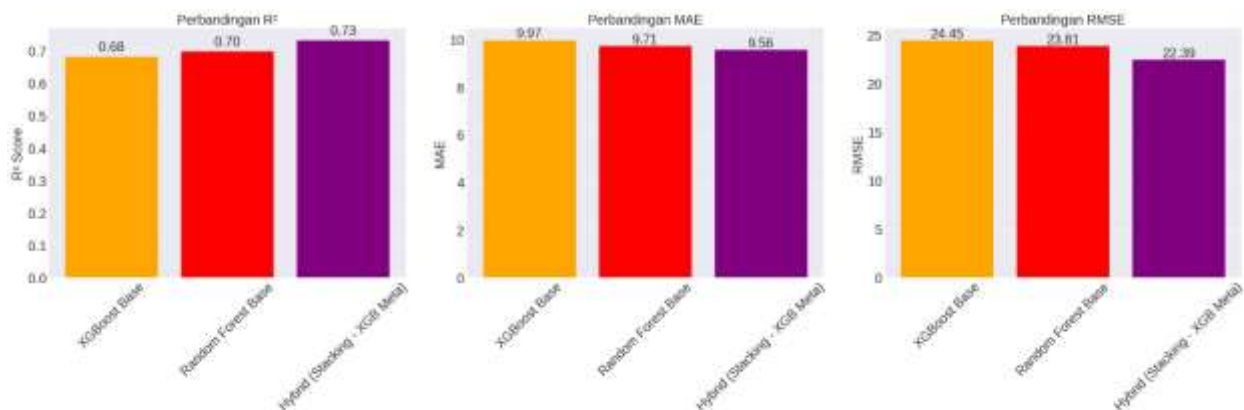


Gambar 3. Mutual Information Scores antara fitur dan target.

Temuan ini konsisten dengan penelitian sebelumnya bahwa prediksi PM2.5 jangka pendek sangat bergantung pada persistensi temporal konsentrasinya, terutama nilai PM2.5 periode sebelumnya. Zamani Joharestani et al. [6] menunjukkan bahwa PM2.5 lag menjadi prediktor kunci dalam Random Forest dan XGBoost, bahkan lebih informatif daripada variabel eksternal seperti AOD. Secara fisis, hal ini masuk akal karena PM2.5 bersifat akumulatif dan tak langsung hilang dalam waktu singkat, terutama saat stagnasi udara. Di Indonesia, Cholianawati et al. [16] mengaitkan variasi PM2.5 di Pontianak dengan pola diurnal, musiman, dan meteorologi multivariat, sementara Istiana et al. [17] menemukan hubungan musiman dan nonlinear dengan parameter cuaca. Karena itu, lemahnya korelasi linear suhu, kelembapan, dan tekanan udara dalam studi ini bukan berarti faktor meteorologi tidak berperan, melainkan pengaruhnya bersifat tidak langsung, nonlinear, atau termanifestasi lewat interaksi dengan pola temporal PM2.5. Hal ini menguatkan relevansi penggunaan fitur lag, rolling mean, dan ensemble machine learning untuk menangkap persistensi temporal dan hubungan nonlinear dalam prediksi PM2.5 per jam di Pontianak.

3.3. Evaluasi kinerja model

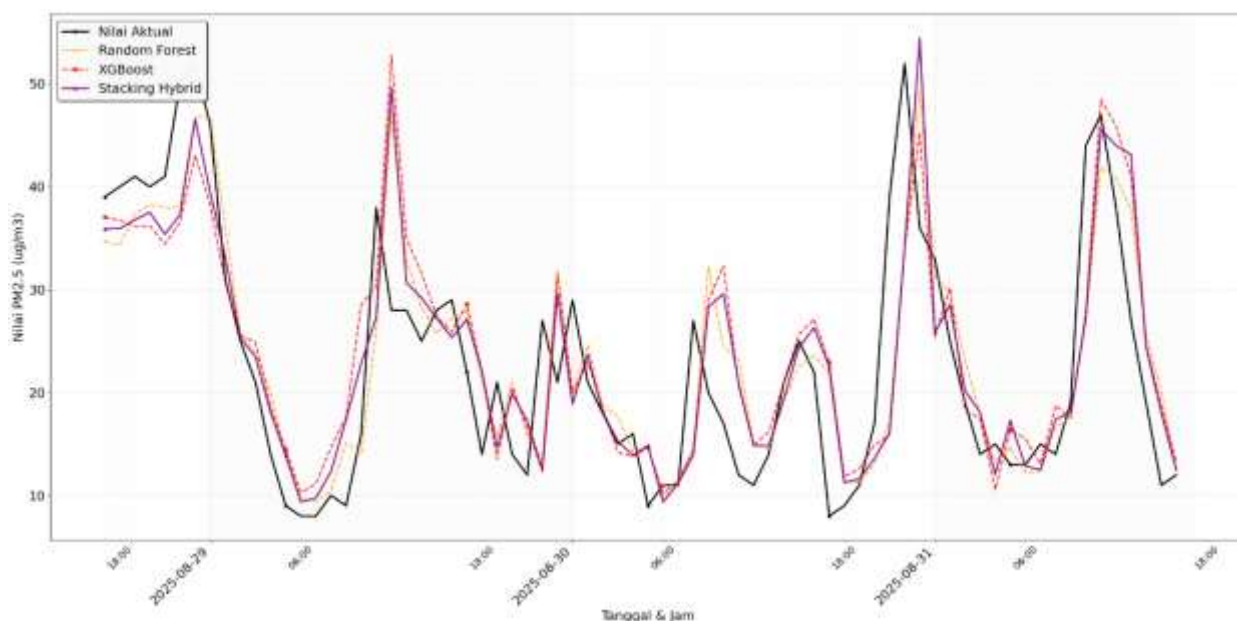
Hasil perbandingan pada Gambar 4 menunjukkan bahwa model *Hybrid Ensemble (Stacking: RF + XGBoost Meta)* memberikan kinerja terbaik dengan nilai $R^2 = 0,72$, MAE = 9,71 $\mu\text{g}/\text{m}^3$, dan RMSE = 22,79 $\mu\text{g}/\text{m}^3$. Nilai ini lebih tinggi dibandingkan model tunggal *Random Forest* ($R^2 = 0,69$) dan *XGBoost* ($R^2 = 0,65$), dengan tingkat kesalahan yang lebih rendah baik pada MAE maupun RMSE. Peningkatan nilai R^2 sebesar sekitar 4–7% dibandingkan model individual menunjukkan bahwa kombinasi dua algoritma berbasis *ensemble learning* mampu mengoptimalkan kekuatan masing-masing: Random Forest yang kuat terhadap varians data dan XGBoost yang unggul dalam menangani bias serta fitur *non-linear* [18].



Gambar 4. Perbandingan kinerja model (R^2 , MAE, RMSE).

Dari sisi stabilitas kesalahan, model *ensemble* juga menunjukkan peningkatan yang konsisten. Nilai MAE menurun dari 10,04 (*XGBoost*) dan 9,89 (*Random Forest*) menjadi 9,71 pada model hybrid, sedangkan RMSE turun cukup signifikan dari 25,48 menjadi 22,79. Penurunan RMSE ini menandakan kemampuan model *ensemble* dalam mengurangi kesalahan besar (*outlier errors*) yang sering muncul pada periode lonjakan ekstrem PM2.5. Dengan demikian, kombinasi model berbasis *stacking* memberikan estimasi yang lebih halus, adaptif terhadap variasi temporal, dan lebih tangguh dalam menangani dinamika data polutan yang fluktuatif.

Visualisasi perbandingan hasil prediksi terhadap data aktual (Gambar 5) memperlihatkan konsistensi model *ensemble* dalam mengikuti pola variasi PM2.5 aktual. Baik puncak maupun lembah konsentrasi berhasil direpresentasikan dengan lebih baik dibandingkan model individual. Pada beberapa periode lonjakan, model *XGBoost* cenderung sedikit *underestimate*, sementara *Random Forest* kadang *overestimate* nilai puncak. Sebaliknya, model *ensemble* mampu menyeimbangkan kedua kecenderungan tersebut dengan mengikuti kontur aktual secara lebih akurat. Pola ini mengindikasikan bahwa pendekatan *stacking ensemble* efektif dalam menangkap hubungan kompleks antara komponen historis PM2.5 dan variabel meteorologi tanpa kehilangan stabilitas generalisasi model.



Gambar 5. Perbandingan prediksi PM2.5 aktual dan hasil model.

Secara keseluruhan, hasil pemodelan ini membuktikan bahwa pendekatan *Hybrid Ensemble (Stacking)* lebih unggul dalam memprediksi konsentrasi setiap jam PM2.5 di Pontianak dibandingkan model tunggal. Kombinasi dua algoritma berbasis *decision tree ensemble* memberikan keseimbangan yang baik antara akurasi dan ketahanan terhadap gangguan (*noise*) data. Temuan ini mendukung penerapan metode *machine learning ensemble* dalam sistem peringatan dini kualitas udara di wilayah tropis, di mana kondisi atmosfer cenderung dinamis dan hubungan antarvariabel bersifat nonlinear.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan model *ensemble machine learning* yang mampu meningkatkan akurasi prediksi konsentrasi per jam PM2.5 di Kota Pontianak. Berdasarkan data observasi selama Agustus 2024–Agustus 2025, konsentrasi PM2.5 menunjukkan pola musiman yang jelas, dengan peningkatan signifikan pada musim kemarau dan penurunan pada musim hujan, yang diduga dipengaruhi oleh kondisi meteorologi dan aktivitas kebakaran lahan di wilayah sekitarnya. Analisis korelasi dan *mutual information* mengindikasikan bahwa fitur temporal seperti *lagged PM2.5* dan *rolling average* memiliki pengaruh paling dominan terhadap konsentrasi PM2.5, sedangkan variabel meteorologi memberikan kontribusi yang relatif kecil. Dari sisi performa model, kombinasi *Random Forest* dan *XGBoost* dalam skema *Hybrid Stacking* menghasilkan akurasi terbaik ($R^2 = 0,72$; MAE = 9,71 $\mu\text{g}/\text{m}^3$; RMSE = 22,79 $\mu\text{g}/\text{m}^3$), lebih unggul dibandingkan model individual. Hasil ini menegaskan bahwa model *ensemble* berbasis pembelajaran mesin efektif untuk memprediksi fluktuasi PM2.5 pada lingkungan tropis yang dinamis serta berpotensi diaplikasikan dalam sistem peringatan dini kualitas udara di Pontianak dan kawasan serupa.

Ucapan Terima Kasih

Penelitian ini dibiayai oleh Dana DIPA Universitas Tanjungpura Nomor SP DIPA-139.03.2.693380/2026 1 Desember 2025. Penulis menyampaikan apresiasi dan terima kasih kepada Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) serta Dekan FMIPA Universitas Tanjungpura atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini.

Pustaka

- [1] Nazarenko Y, Pal D and Ariya P A 2021 Air quality standards for the concentration of particulate matter 2.5, global descriptive analysis *Bull. World Health Organ.* **99** 125–137
- [2] Cohen A J, Brauer M, Burnett R, Anderson H R, Frostad J, Estep K *et al.* 2017 Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015 *Lancet* **389** 1907–1918
- [3] Giannadaki D, Lelieveld J and Pozzer A 2016 Implementing the US air quality standard for PM_{2.5} worldwide can prevent millions of premature deaths per year *Environ. Health* **15** 88
- [4] Umah R and Gusmira E 2024 Dampak pencemaran udara terhadap kesehatan masyarakat di perkotaan *Profit: Jurnal Manajemen, Bisnis dan Akuntansi* **3** 103–112
- [5] BMKG 2024 *Laporan Kualitas Udara dan Analisis Cuaca untuk Wilayah Kalimantan Barat* (Jakarta: BMKG)
- [6] Zamani Joharestani M, Cao C, Ni X, Bashir B and Talebiesfandarani S 2019 PM_{2.5} prediction based on Random Forest, XGBoost, and deep learning using multisource remote sensing data *Atmosphere* **10** 373
- [7] Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* pp 785–794
- [8] Zhou Z-H 2025 *Ensemble Methods: Foundations and Algorithms* (Boca Raton: CRC Press)
- [9] Peng H, Long F and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–1238
- [10] Willmott C J and Matsuura K 2005 Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance *Clim. Res.* **30** 79–82
- [11] Chai T and Draxler R R 2014 Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature *Geosci. Model Dev.* **7** 1247–1250
- [12] Chen M, Liu J, Chu B, Zhao D, Li R, Chen T *et al.* 2025 Research on the influencing factors of PM_{2.5} in China at different spatial scales based on machine learning algorithm *Environ. Int.* **200**
- [13] Gao Z, Do K, Li Z, Jiang X, Maji K J, Ivey C E and Russell A G 2024 Predicting PM_{2.5} levels and exceedance days using machine learning methods *Atmos. Environ.* **323** 120396
- [14] IQAir 2025 AirVisual Outdoor 2 PM Available from: <https://www.iqair.com/id/products/air-quality-monitors/airvisual-outdoor-2-pm> [accessed 18 July 2025]
- [15] Ais U P and Sanjaya S R 2025 Perbandingan algoritma Random Forest, XGBoost, dan Logistic Regression untuk prediksi risiko kekambuhan kanker tiroid *Edumatic: J. Pendidik. Inform.* **9** 236–245
- [16] Cholianawati N, Sinatra T, Nugroho G A, Permadi D A, Indrawati A, Halimurrahman *et al.* 2024 Diurnal and daily variations of PM_{2.5} and its multiple-wavelet coherence with meteorological variables in Indonesia *Aerosol Air Qual. Res.* **24** 230158
- [17] Istiana T, Kurniawan B, Soekirno S, Nahas A, Wihono A, Nuryanto D E *et al.* 2023 Causality analysis of air quality and meteorological parameters for PM_{2.5} characteristics determination: evidence from Jakarta *Aerosol Air Qual. Res.* **23** 230014
- [18] Awan Irawan H 2025 Deteksi fraud kartu kredit dengan Logistic Regression, Random Forest, dan Gradient Boosting *J. Inform. Terpadu* **11** 92–97.