

**EVALUASI KINERJA XGBOOST (*EXTREME GRADIENT BOOSTING*)
DAN *RANDOM FOREST* UNTUK ANALISA KUALITAS UDARA KOTA
BANDAR LAMPUNG**

(Skripsi)

Oleh

RITA IRMA YANI

2217051047



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

**EVALUASI KINERJA XGBOOST (*EXTREME GRADIENT BOOSTING*)
DAN *RANDOM FOREST* UNTUK ANALISA KUALITAS UDARA KOTA
BANDAR LAMPUNG**

Oleh

RITA IRMA YANI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA ILMU KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

ABSTRAK

EVALUASI KINERJA XGBOOST (*EXTREME GRADIENT BOOSTING*) DAN *RANDOM FOREST* UNTUK ANALISA KUALITAS UDARA KOTA BANDAR LAMPUNG

Oleh

RITA IRMA YANI

Pemantauan kualitas udara merupakan hal yang esensial dalam penilaian kesehatan lingkungan, khususnya di wilayah perkotaan yang berkembang pesat. Penelitian ini mengevaluasi dan membandingkan kinerja *Random Forest Regression* dan *Extreme Gradient Boosting (XGBoost) Regression* dalam menganalisis kualitas udara di Kota Bandar Lampung, Indonesia, menggunakan data harian dari stasiun *Automatic Air Quality Monitoring System (AQMS) Sukarame* periode 2023-2025. Dataset terdiri dari 1.073 observasi harian dengan sembilan variabel input PM₁₀, PM_{2,5}, SO₂, CO, O₃, NO₂, kelembaban, kecepatan angin, dan suhu serta Indeks Standar Pencemar Udara (ISPU) sebagai variabel target. Praproses data mencakup interpolasi linear berbasis waktu untuk nilai yang hilang dan perhitungan ISPU berdasarkan Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor 14 Tahun 2020. Pembagian data latih-uji 80:20 dengan pengacakan diterapkan setelah mengevaluasi tiga proporsi pembagian (70:30, 80:20, 85:15). Evaluasi kinerja menggunakan MAPE, RMSE, MAE, dan R². *Random Forest* mencapai MAPE = 1,178%, RMSE = 1,995, MAE = 0,744, dan R² = 0,990, sedangkan XGBoost mencapai MAPE = 1,570%, RMSE = 2,075, MAE = 0,969, dan R² = 0,989. Analisis *feature importance* dari kedua model secara konsisten mengidentifikasi PM_{2,5} sebagai variabel paling dominan, diikuti oleh SO₂, NO₂, dan O₃, sementara variabel meteorologi berkontribusi secara minimal. *Random Forest* menunjukkan kinerja keseluruhan yang sedikit lebih unggul, meskipun kedua model terbukti sangat efektif untuk analisis kualitas udara di Kota Bandar Lampung.

Kata kunci: Analisis Kualitas Udara, *Random Forest*, XGBoost, ISPU, Pembelajaran Mesin.

ABSTRACT

PERFORMANCE EVALUATION OF RANDOM FOREST AND XGBOOST FOR AIR QUALITY ANALYSIS IN BANDAR LAMPUNG

By

RITA IRMA YANI

Air quality monitoring is essential for environmental health assessment, particularly in rapidly growing urban areas. This study evaluates and compares the performance of *Random Forest Regression* and *Extreme Gradient Boosting (XGBoost) Regression* in analyzing air quality in Bandar Lampung City, Indonesia, using daily data from the *Automatic Air Quality Monitoring System (AQMS)* Sukarame station for the period 2023-2025. The dataset comprises 1,073 daily observations with nine input variables PM₁₀, PM_{2.5}, SO₂, CO, O₃, NO₂, humidity, wind speed, and temperature and the Air Pollutant Standard Index (ISPU) as the target variable. Preprocessing included time-based linear interpolation for missing values and ISPU calculation based on Indonesian Ministry of Environment Regulation No. 14/2020. An 80:20 train-test split with random shuffling was applied after evaluating three split proportions (70:30, 80:20, 85:15). Performance was assessed using MAPE, RMSE, MAE, and R². *Random Forest* achieved MAPE = 1.178%, RMSE = 1.995, MAE = 0.744, and R² = 0.990, while XGBoost achieved MAPE = 1.570%, RMSE = 2.075, MAE = 0.969, and R² = 0.989. Feature importance analysis from both models consistently identified PM_{2.5} as the most dominant variable, followed by SO₂, NO₂, and O₃, while meteorological variables contributed minimally. *Random Forest* demonstrated slightly superior overall performance, though both models proved highly effective for air quality analysis in Bandar Lampung City.

Keywords: Air Quality Analysis, Random Forest, XGBoost, ISPU, Machine Learning

Judul Skripsi

EVALUASI KINERJA XGBOOST (*EXTREME GRADIENT BOOSTING*) DAN *RANDOM FOREST* UNTUK ANALISA KUALITAS UDARA KOTA BANDAR LAMPUNG

Nama Mahasiswa

Rita Irma Yani

Nomor Pokok Mahasiswa

2217051047

Program Studi

Ilmu Komputer

Fakultas

Matematika dan Ilmu Pengetahuan Alam



1. Komisi Pembimbing

Prof. Admi Syarif, Ph.D

NIP. 19670103 199203 1 003

Ridho Sholehurrohman, M.Mat

NIP. 232111970128101

2. Mengetahui

Ketua Jurusan Ilmu Komputer

Ketua Program Studi

Dwi Sakethi, S.Si., M.Kom

NIP. 19680611 199802 1 001

Tristiyanto, M.I.S., Ph.D

NIP. 198104142005011001

MENGESAHKAN

1. Tim Penguji

Ketua : **Prof. Admi Syarif, Ph.D**

Sekretaris : **Ridho Sholehurrohman, M. Mat**

Penguji Utama : **Dr. Erdi Suroso, S.T.P., M.T.A., C.EIA**

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.

NIP-19711001 200501 1 002



Tanggal Lulus Ujian Skripsi : 04 Juni 2026

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rita Irma Yani

NPM : 2217051047

Dengan ini menyatakan bahwa skripsi yang berjudul " Evaluasi Kinerja XGBoost (*Extreme Gradient Boosting*) Dan *Random Forest* Untuk Analisa Kualitas Udara Kota Bandar Lampung" Merupakan Hasil Karya Saya Sendiri Dan Bukan Karya Orang Lain. Semua tulisan yang tertulis pada skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari ditemukan bukti bahwa skripsi saya merupakan hasil penjiplakan atau dibuat oleh orang lain maka saya bersedia menerima sanksi sesuai hukum yang berlaku.

Bandar Lampung, 12 Juni 2026



Rita Irma Yani

NPM. 2217051047

RIWAYAT HIDUP



Penulis dilahirkan di Kartasari, Tulang Bawang Barat pada tanggal 10 Januari 2003, sebagai anak pertama dari dua bersaudara, dari bapak Sukari dan ibu Yuliani. Pendidikan yang sudah di tempuh oleh penulis diantaranya, menyelesaikan Pendidikan Taman Kanak-kanan (TK) Dharmawanita Kartasari, Tulang Bawang Barat dan lulus pada tahun 2010. Kemudian penulis melanjutkan Pendidikan Sekolah Dasar di SDN 1 Kartasari pada tahun 2010 – 2016, lalu setelahnya melanjutkan Sekolah Menengah Pertama di SMPN 2 Tulang Bawang Udik pada tahun 2016 – 2019 dan Sekolah Menengah Atas di SMAN 1 Tumijajar pada tahun 2019 – 2022.

Pada Tahun 2022 penulis terdaftar sebagai mahasiswa Program Studi S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri). Selama menempuh studi di perguruan tinggi, penulis terlibat dalam beberapa kegiatan sebagai berikut.

1. Finalis Musabaqah Tilawatil Qur'an Mahasiswa (MTQM) Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung Tahun 2022 dan 2023.
2. Anggota Bidang Kaderisasi Unit Kegiatan Mahasiswa Rohani Islam (Rois) Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung Tahun 2023.
3. Anggota Bidang Keilmuan Himpunan Mahasiswa Ilmu Komputer Universitas Lampung Tahun 2023.

4. Peserta Musabaqah Tilawatil Qur'an Mahasiswa Nasional (MTQMN) yang diselenggarakan oleh Universitas Negeri Yogyakarta (UNY) Tahun 2023.
5. Sekretaris Bidang Kaderisasi Unit Kegiatan Mahasiswa Rohani Islam (Rois) Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung Tahun 2024.
6. Asisten dosen pada Program Studi S1 Ilmu Komputer Universitas Lampung untuk Mata Kuliah Matematika Diskrit serta Komunikasi Data dan Jaringan Komputer Tahun 2024, serta Mata Kuliah Statistika dan Probabilitas Tahun 2025.
7. Mengikuti Kegiatan Kerja Praktik di PT Setia Karya Transport, sebuah unit kerja dari *Great Giant Foods* (GGF) yang bergerak di bidang jasa transportasi Tahun 2025.
8. Peserta Kuliah Kerja Nyata (KKN) di Desa Triharjo Kecamatan Merbau Mataram, Kabupaten Lampung Selatan Provinsi Lampung Periode 1 Tahun 2025.
9. Mengikuti organisasi di luar kampus yaitu menjadi Ketua Bidang Kajian dan keagamaan Pergerakan Mahasiswa Islam Indonesia (PMII), Rayon Matematika dan Ilmu Pengetahuan Alam Universitas Lampung Tahun 2024.

MOTTO

مَنْ جَدَّ وَجَدَ

“Barang siapa bersungguh-sungguh, maka ia akan berhasil”

مَنْ صَبَرَ ظَفِرَ

“Barang siapa yang bersabar, ia akan beruntung (atau menang)”

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”

(Q.S. Al-Baqarah Ayat 286)

“Kalau Tuhan sudah bawa kamu sejauh ini, pasti bukan untuk gagal, tapi karena ada hal baik yang sudah disediakan”

Jika kamu tidak sanggup menahan lelahnya belajar, maka kamu harus sanggup menahan perihnya kebodohan”

(Imam Syafi’i)

“To get something as beautiful as the sky, we also need to do the paths of the sky”

(Anonymous)

PERSEMBAHAN

Alhamdulillah Robbil 'alamiin

Segala puji dan syukur saya panjatkan kepada Allah Subhanahu wa Ta'ala atas segala kemudahan, rahmat, karunia serta hidayahnya sehingga skripsi ini dapat terselesaikan. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad Shallallahu Alaihi Wasallam, suri teladan bagi seluruh umat manusia. karya ini dipersembahkan sebagai tanda bakti, cinta, dan kasih sayang kepada orang-orang istimewa yang senantiasa hadir dalam perjalanan hidup penulis, kepada:

Kedua Orang Tua, Adik dan Nenek Tercinta

Untuk kedua orang tua, adik, dan nenek tercinta yang senantiasa menjadi kekuatan terbesar dalam setiap langkah perjalanan penulis. Ibu, terima kasih telah menjadi tempat paling teduh dalam setiap doa dan kasih sayangmu sebab tidak ada hadiah paling berharga selain doa seorang ibu yang senantiasa menyertai langkah anaknya. Ayah, terima kasih telah mengajarkan penulis untuk berani melangkah, kuat menghadapi keadaan, dan terus maju tanpa mudah menyerah. Adik tercinta, terima kasih telah menjadi salah satu alasan penulis untuk terus bertahan hingga akhir, semoga kelak menjadi pribadi yang membanggakan dan meraih segala cita-cita yang diimpikan. Nenek tercinta, terima kasih atas doa tulus yang selalu menyertai langkah penulis sejak kecil hingga saat ini, membawa harapan terbaik untuk pendidikan, masa depan, dan kebahagiaan penulis.

Seluruh Keluarga Besar Ilmu Komputer

Yang selalu membantu dan memberikan dukungan.

Almamater yang Kubanggakan, Universitas Lampung

Tempat bernaung dan mengemban semua ilmu.

SANWACANA

Puji Syukur penulis ucapkan kehadiran Allah Subhanahu wa Ta'ala, karena atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Evaluasi Kinerja XGBoost (*Extreme Gradient Boosting*) Dan *Random Forest* Untuk Analisa Kualitas Udara Kota Bandar Lampung” telah diselesaikan. Tidak lupa shalawat dan salam yang senantiasa dicurahkan kepada baginda Nabi Muhammad Shallallahu Alaihi Wasallam yang menjadi suri teladan bagi seluruh umat manusia.

Dengan penuh rasa syukur dan kerendahan hati, pada kesempatan ini penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini, yaitu:

1. Allah Subhanahu wa Ta'ala yang telah melimpahkan berkah, hidayah, rahmat, serta karunia-Nya sehingga proses penyelesaian skripsi ini dapat berjalan dengan lancar.
2. Bapak Sukari dan Ibu Yuliani selaku kedua orang tua tercinta yang senantiasa menjadi kekuatan terbesar, tak pernah berhenti mendoakan, memotivasi, serta memberikan dukungan dan bantuan yang tidak terhitung nilainya, adik tersayang Alfi Ma'ruf yang selalu menjadi alasan untuk terus berjuang, serta nenek tercinta Rasmimi yang doa tulusnya tak pernah putus mengiringi langkah penulis.
3. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
4. Bapak Dwi Sakethi, S.Si., M.Kom., selaku Ketua Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Ibu Yunda Heningtyas, M.Kom., selaku Sekretaris Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

6. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D., selaku Ketua Program Studi S1 Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Bapak Prof. Admi Syarif, Ph.D., selaku Dosen Pembimbing Utama yang selalu membimbing, memberikan arahan, motivasi, masukan serta saran, dalam penyelesaian skripsi ini.
8. Bapak Ridho Sholehurrohman, M.Mat., selaku Dosen Pembimbing kedua yang selalu membimbing, memberikan arahan, masukan dan saran dalam penyelesaian skripsi ini.
9. Bapak Dr. Erdi Suroso, S.T.P., M.T.A., C.EIA., selaku Dosen Pembahas yang telah memberikan saran dan masukan yang bermanfaat dalam perbaikan skripsi ini.
10. Bapak Bambang Hermanto, S.Kom., M.Cs., selaku Dosen Pembimbing Akademik yang telah memberikan motivasi dan saran kepada penulis.
11. Ibu Ade Nora, staf Jurusan Ilmu Komputer yang telah membantu berbagai urusan administrasi selama masa perkuliahan hingga penyelesaian skripsi ini.
12. Bapak/Ibu Dosen serta seluruh staf Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan membantu dalam berbagai keperluan akademik maupun administratif.
13. Dinas Lingkungan Hidup Kota Bandar Lampung beserta seluruh staf yang telah memberikan izin dan menyediakan data Indeks Standar Pencemar Udara (ISPU) yang penulis perlukan dalam proses penyelesaian skripsi ini.
14. Kakak-kakak tingkat terdekat yang tak pernah lelah membantu meski sering penulis repotkan, selalu hadir dengan masukan yang meluruskan, serta kata-kata singkat namun penuh makna. Partner mumet yang turut membantu dalam proses penyelesaian skripsi ini, guru ngaji, teman-teman Rois, serta seluruh keluarga besar yang senantiasa memberikan doa, nasihat, dan semangat.
15. Citra Fitria Sari, sepupu sekaligus motivator sejak kecil yang turut berkontribusi dalam penelitian ini. Jusie Nora Afrida, sahabat sejak SD hingga kini yang terasa seperti keluarga, terima kasih sudah selalu ada. Serta orang-orang terdekat yang turut terlibat dan memberikan dukungan dalam proses ini.

16. "Tiga Serangkai", Elsa Dwi Agresty dan Yunia Saputri, sahabat terbaik sejak awal perkuliahan hingga penyelesaian skripsi ini. Meski kini kesibukan telah membawa kita pada jalan masing-masing, kalian tetap menjadi bagian paling berarti dalam perjalanan penulis.
17. Zainab Aqilah, partner seperjuangan sejak awal penulisan skripsi hingga penyelesaiannya, terima kasih telah saling membantu dan menguatkan dalam setiap prosesnya. Meski tak selalu berjalan mulus, kebersamaan ini tetap menjadi bagian yang paling berkesan dalam perjalanan penulis.
18. Evi, sahabat sekaligus teman kos, tempat bercerita dan berbagi keluh kesah, terima kasih sudah selalu ada di setiap momen. Nurjayatri, Ulfa, dan Rica, yang kehadirannya selalu memberi warna tersendiri bagi penulis. Serta Sirkel Mitha dan teman-teman lain yang tidak dapat disebutkan satu per satu terima kasih telah menjadi bagian dari perjalanan ini.
19. Teman-teman Ilmu Komputer Angkatan 2022 yang menjadi keluarga besar Jurusan Ilmu Komputer selama menjalankan masa studi.
20. Terakhir, untuk diri sendiri yang tak pernah berhenti bermimpi, Rita Irma Yani anak perempuan pertama, yang kadang kurang percaya diri namun tak pernah berhenti mencoba. Terima kasih telah kuat dalam berjuang, mampu melewati setiap rintangan, dan tak pernah memutuskan untuk menyerah sesulit apapun prosesnya. Semua ini bukan semata hasil usaha sendiri, melainkan karena Allah Subhanahu wa Ta'ala yang memampukan. Berbahagialah dan teruslah tumbuh dimanapun berada untuk mencapai impian.

Penulis menyadari skripsi ini masih terdapat banyak kekurangan karena keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan. Semoga skripsi ini dapat membawa manfaat dan keberkahan bagi seluruh civitas Jurusan Ilmu Komputer Universitas Lampung.

Bandar Lampung, 12 Juni 2026



Rita Irma Yani
NPM. 2217051047

DAFTAR ISI

	Halaman
DAFTAR ISI	i
DAFTAR GAMBAR	iv
DAFTAR TABEL	vi
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
II. TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu.....	7
2.2 Landasan Teori.....	12
2.2.1 Kota Bandar Lampung.....	12
2.2.2 Kualitas Udara.....	18
2.2.3 Polutan.....	20
2.2.4 <i>Interpolasi Linear</i>	26
2.2.5 Indeks Standar Pencemar Udara (ISPU).....	27
2.2.6 <i>Artificial Intelligence (AI)</i>	28
2.2.7 <i>Regression</i>	31
2.2.8 <i>Random Forest</i>	31
2.2.9 <i>Extreme Gradient Boosting (XGBoost)</i>	34
2.2.10 Evaluasi Model.....	38
III. METODOLOGI PENELITIAN	42
3.1 Tempat dan Waktu.....	42
3.1.1 Tempat Penelitian.....	42
3.1.2 Waktu Penelitian.....	42
3.2 Alat Pendukung.....	43

3.2.1 Perangkat Keras	43
3.2.2 Perangkat Lunak	43
3.3 Tahapan Penelitian	43
3.3.1 Studi Literatur	45
3.3.2 Pengumpulan Data	46
3.3.3 Eksplorasi Data	47
3.3.4 <i>Preprocessing</i>	50
3.3.5 <i>Splitting Data</i>	54
3.3.6 Evaluasi Kinerja Model	55
3.3.7 Analisis <i>Feature Importance</i>	56
3.3.8 Perbandingan Hasil Pengujian Model	57
IV. HASIL DAN PEMBAHASAN	58
4.1 Pengumpulan Data	58
4.2 Eksplorasi Data	59
4.2.1 Analisis Deskriptif	59
4.2.2 <i>Missing Value</i>	62
4.2.3 Grafik Tren Data	62
4.2.4 Histogram Distribusi Data	69
4.2.5 Korelasi Antar Variabel	75
4.2.6 Deteksi <i>Outlier</i>	78
4.3 <i>Preprocessing</i> Data	80
4.3.1 <i>Handling Missing Value</i>	80
4.3.2 Perhitungan Indeks Standar Pencemar Udara (ISPU)	83
4.3.3 Visualisasi Grafik ISPU	88
4.3.4 <i>Feature Selection</i>	89
4.4 <i>Splitting Data</i>	91
4.4.1 <i>Data Training</i>	93
4.4.2 <i>Data Testing</i>	93
4.5 Evaluasi Kinerja Model	94
4.5.1 <i>Random Forest Regression</i>	94
4.5.2 <i>XGBoost Regression</i>	95
4.6 Analisis <i>Feature Importance</i>	96
4.6.1 Analisis <i>Feature Importance Random Forest Regression</i>	96
4.6.2 Analisis <i>Feature Importance XGBoost Regression</i>	98
4.7 Perbandingan Hasil Pengujian <i>Random Forest</i> dan <i>XGboost</i>	100

V. SIMPULAN DAN SARAN.....	104
5.1 Simpulan	104
5.2 Saran.....	105
DAFTAR PUSTAKA.....	106
LAMPIRAN.....	111

DAFTAR GAMBAR

Gambar	Halaman
1. Kota Bandar Lampung.....	13
2. Peta Provinsi Lampung.....	15
3. Peta Kota Bandar Lampung.....	16
4. Peta Titik Lokasi Stasiun Pemantauan Kualitas Udara (AQMS) di Kota Bandar Lampung.....	17
5. Kategori <i>Machine Learning</i> Berdasarkan Cara Belajarnya.....	29
6. Alur Algoritma <i>Random Forest</i>	33
7. Alur Algoritma <i>XGBoost</i>	37
8. Alur Tahapan Penelitian.....	44
9. Grafik Tren Konsentrasi PM ₁₀ (µg/m ³) Kota Bandar Lampung.....	63
10. Grafik Tren Konsentrasi PM _{2,5} (µg/m ³) Kota Bandar Lampung.....	64
11. Grafik Tren Konsentrasi SO ₂ (µg/m ³) Kota Bandar Lampung.....	64
12. Grafik Tren Konsentrasi CO (µg/m ³) Kota Bandar Lampung.....	65
13. Grafik Tren Konsentrasi O ₃ (µg/m ³) Kota Bandar Lampung.....	66
14. Grafik Tren Konsentrasi NO ₂ (µg/m ³) Kota Bandar Lampung.....	67
15. Grafik Tren Kelembaban (%) Kota Bandar Lampung.....	67
16. Grafik Tren Kecepatan Angin (m/s) Kota Bandar Lampung.....	68
17. Grafik Tren Suhu (°C) Kota Bandar Lampung.....	69
18. Histogram Distribusi Konsentrasi PM ₁₀ (µg/m ³).....	70
19. Histogram Distribusi Konsentrasi PM _{2,5} (µg/m ³).....	70
20. Histogram Distribusi Konsentrasi SO ₂ (µg/m ³).....	71
21. Histogram Distribusi Konsentrasi CO (µg/m ³).....	72
22. Histogram Distribusi Konsentrasi O ₃ (µg/m ³).....	72
23. Histogram Distribusi Konsentrasi NO ₂ (µg/m ³).....	73

24. Histogram Distribusi Kelembaban (%)	74
25. Histogram Distribusi Kecepatan Angin (m/s).....	74
26. Histogram Distribusi Suhu (°C).....	75
27. Heatmap Korelasi Antar Variabel	76
28. Deteksi <i>Outlier</i>	78
29. Visualisasi Grafik ISPU Kualitas Udara Kota Bandar Lampung	88
30. <i>Feature Importance Random Forest Regression</i>	97
31. <i>Feature Importance XGboost Regression</i>	99
32. Perbandingan Hasil Pengujian <i>Random Forest</i> dan <i>XGboost</i>	103

DAFTAR TABEL

Tabel	Halaman
1. Penelitian Terdahulu.....	7
2. Kriteria Tingkat Akurasi berdasarkan Nilai MAPE	39
3. Waktu Penelitian	42
4. Analisis Deskriptif Data Kualitas Udara Kota Bandar Lampung.....	60
5. Jumlah <i>Missing Value</i> pada Data Kualitas Udara Kota Bandar Lampung.....	62
6. Ilustrasi <i>Missing Value</i> Sebelum <i>Interpolasi Linear</i>	81
7. Ilustrasi Data Setelah Penanganan <i>Missing Value</i> Menggunakan <i>Interpolasi Linear</i>	83
8. Konversi Nilai Konsentrasi Parameter ISPU	84
9. Data Konsentrasi Parameter Polutan dalam kurun waktu 24 jam Sebelum Perhitungan ISPU.....	85
10. Hasil Perhitungan ISPU Parameter Polutan.....	87
11. Kategori Angka Rentang ISPU	88
12. Variabel Fitur (X) Parameter Polutan dan Meteorologi	90
13. Variabel Target (y) Nilai ISPU	90
14. Perbandingan Kinerja Model Berdasarkan Proporsi Pembagian Data	91
15. Hasil Evaluasi <i>Random Forest Regression</i>	94
16. Hasil Evaluasi <i>Xgboost Regression</i>	95
17. Nilai <i>Importance Random Forest Regression</i>	96
18. Nilai <i>Feature Importance XGBoost</i>	98
19. Hasil Evaluasi Model pada Penelitian Terdahulu dan Penelitian Kota Bandar Lampung	101

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Kualitas udara merupakan salah satu indikator penting dalam menilai kondisi kesehatan lingkungan di suatu wilayah. Udara yang bersih dan sehat tidak hanya mendukung keberlangsungan ekosistem, tetapi juga menjadi faktor utama dalam menjaga kesehatan masyarakat. Udara yang terkontaminasi oleh berbagai polutan dapat memberikan dampak negatif bagi kesehatan manusia, seperti gangguan pernapasan, penyakit kardiovaskular, hingga menurunnya kualitas hidup secara umum (Rahmawati et al., 2024).

Selain berdampak pada kesehatan, kualitas udara yang buruk juga dapat mempengaruhi sektor ekonomi, seperti meningkatnya biaya kesehatan dan menurunnya produktivitas masyarakat. Penurunan kualitas udara ini umumnya dipicu oleh aktivitas manusia, terutama pertumbuhan industri dan meningkatnya jumlah kendaraan bermotor yang menghasilkan polutan seperti debu, karbon monoksida, dan partikel lain yang dapat mengganggu kesehatan. Oleh karena itu, pemantauan dan analisis kualitas udara perlu dilakukan secara berkala untuk mengetahui tingkat pencemaran dan potensi risikonya terhadap kesehatan masyarakat (Endra & Saputra, 2022).

Salah satu indikator yang digunakan untuk menggambarkan kondisi kualitas udara adalah Indeks Standar Pencemar Udara (ISPU). ISPU merupakan suatu angka yang menggambarkan kondisi mutu udara ambien di suatu lokasi berdasarkan konsentrasi beberapa parameter pencemar utama seperti PM_{10} , $PM_{2,5}$, SO_2 , CO , O_3 , dan NO_2 . Nilai ISPU diperoleh melalui proses pemantauan kualitas udara menggunakan stasiun pemantauan udara yang beroperasi secara terus menerus. Nilai tersebut kemudian digunakan sebagai indikator untuk

menentukan kategori kualitas udara, seperti baik, sedang, tidak sehat, sangat tidak sehat, hingga berbahaya (Manik et al., 2026).

Pemantauan kualitas udara di Kota Bandar Lampung dilakukan menggunakan stasiun *Automatic Air Quality Monitoring System* (AQMS) yang dikelola oleh Dinas Lingkungan Hidup. Stasiun pemantauan tersebut berada di wilayah Sukarame dan merupakan satu-satunya stasiun pemantauan kualitas udara yang tersedia di Kota Bandar Lampung. Data kualitas udara yang dihasilkan oleh stasiun ini memiliki jangkauan pemantauan sekitar ± 5 km yang dipengaruhi oleh arah angin, sehingga data yang diperoleh merepresentasikan (Binbusayyis et al., 2024). kondisi kualitas udara di wilayah Sukarame dan sekitarnya. Data pemantauan tersebut mencakup berbagai parameter polutan udara seperti PM₁₀, PM_{2.5}, SO₂, CO, O₃, dan NO₂ serta didukung oleh data meteorologi seperti kelembaban, kecepatan angin, dan suhu. Data pemantauan ini digunakan untuk melihat pola perubahan kualitas udara dari waktu ke waktu.

Nilai polutan yang berubah-ubah setiap hari menyebabkan data kualitas udara menjadi kompleks dan sulit dianalisis secara manual (Binbusayyis et al., 2024). Kondisi ini juga terlihat pada data kualitas udara di Kota Bandar Lampung. Sebagai pusat aktivitas transportasi, permukiman, dan perdagangan di bagian selatan Sumatera, Kota Bandar Lampung memiliki dinamika kualitas udara yang fluktuatif sehingga diperlukan analisis lebih lanjut untuk memahami pola perubahan kualitas udara di wilayah tersebut.

Dalam melakukan analisis kualitas udara, data pemantauan yang diperoleh dari stasiun AQMS dapat dimanfaatkan untuk mengidentifikasi pola hubungan antara konsentrasi polutan dan kondisi meteorologi. Namun, karena data kualitas udara memiliki karakteristik yang fluktuatif dan dipengaruhi oleh banyak faktor lingkungan, proses analisis menjadi cukup kompleks apabila dilakukan secara manual. Oleh karena itu, diperlukan suatu pendekatan yang dapat membantu menganalisis data kualitas udara secara lebih efektif (Ali et al., 2025). Salah satu pendekatan yang dapat digunakan adalah *machine learning*, yang mampu mempelajari pola hubungan antarvariabel berdasarkan data yang tersedia.

Seiring dengan perkembangan teknologi, *machine learning* banyak digunakan dalam analisis data lingkungan yang memiliki karakteristik kompleks dan fluktuatif. Pendekatan *Machine learning* ini mampu mempelajari pola hubungan antara berbagai variabel dalam dataset sehingga dapat membantu dalam menganalisis perubahan kualitas udara berdasarkan data historis yang tersedia (Ali et al., 2025). Dalam penelitian ini digunakan pendekatan *machine learning* berbasis *ensemble learning* dengan dua algoritma, yaitu *Random Forest* dan *XGBoost (Extreme Gradient Boosting)*. *Random Forest* merupakan algoritma *ensemble learning* yang bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya sehingga menghasilkan model yang lebih stabil. Sementara itu, *XGBoost* merupakan algoritma *boosting* yang membangun model secara bertahap dengan memperbaiki kesalahan pada setiap proses pelatihannya.

Penelitian mengenai prediksi kualitas udara telah dilakukan oleh berbagai peneliti sebelumnya dengan menggunakan pendekatan *machine learning*. Salah satu penelitian dilakukan oleh (Ravindiran et al., 2023) di kota Visakhapatnam, India, yang memanfaatkan 12 jenis polutan serta 10 parameter meteorologi untuk memprediksi AQI (*Air Quality Index*). Penelitian ini membandingkan lima model, yaitu *LightGBM*, *Random Forest*, *CatBoost*, *AdaBoost*, dan *XGBoost*. Dari hasil pengujian, *CatBoost* menjadi model paling unggul dengan nilai R^2 sebesar 0,9998 dan nilai RMSE yang sangat kecil, sehingga menunjukkan performa prediksi yang sangat akurat.

Kemampuan pendekatan *boosting* juga terlihat pada penelitian (Tırınk, 2025) di Provinsi Iğdır, Türkiye. Penelitian tersebut menggunakan data dari tahun 2016 hingga 2024 dengan variabel polutan dan faktor meteorologi. Tiga model dibandingkan, yaitu *XGBoost*, *LightGBM*, dan *Support Vector Machine (SVM)*. Dari hasil pengujian, *XGBoost* memberikan performa terbaik dengan nilai R^2 mencapai 0,999 dan *error* yang rendah. Hasil ini menunjukkan bahwa model *boosting* mampu memberikan hasil prediksi kualitas udara yang sangat baik.

Konsistensi keunggulan model *boosting* juga terlihat pada penelitian yang dilakukan oleh (Naizabayeva et al., 2025) di Kota Almaty, Kazakhstan. Penelitian ini memanfaatkan data polutan dan meteorologi dari 20 stasiun pemantauan udara serta menerapkan metode STL (*Seasonal-Trend decomposition using Loess*) untuk memisahkan pola musiman sebelum pemodelan dilakukan. Dari tiga model yang dibandingkan, yaitu *XGBoost*, *LightGBM*, dan *weighted ensemble*, hasil terbaik diperoleh dari model *ensemble* (gabungan *XGBoost* dan *LightGBM*) dengan nilai R^2 sebesar 0,9825 untuk $PM_{2,5}$ dan 0,8391 untuk CO. Penelitian ini menguatkan bahwa kombinasi *XGBoost* dan *LightGBM* mampu memberikan performa prediksi yang sangat akurat pada data kualitas udara yang memiliki pola musiman yang kuat.

Ketercapaian performa tinggi pada model-model tersebut juga sejalan dengan penelitian (Lei et al., 2023) di Macau, yang membandingkan berbagai algoritma seperti *Random Forest*, *Support Vector Machine (SVM)*, *XGBoost*, *Artificial Neural Network (ANN)*, dan *Multiple Linear Regression (MLR)* untuk memprediksi PM_{10} , $PM_{2,5}$, dan CO. Berdasarkan hasil pengujian tahun 2020, SVM tercatat sebagai model paling akurat untuk PM_{10} dan CO dengan nilai R^2 sebesar 0,94. Sementara itu, untuk $PM_{2,5}$, *Random Forest* menunjukkan performa paling stabil dengan nilai R^2 sebesar 0,88, meskipun nilai tersebut sama dengan yang diperoleh (*Multiple Linear Regression*) MLR. Hasil ini memperkuat bahwa SVM dan *Random Forest* tetap menjadi pilihan algoritma yang efektif dalam memprediksi kualitas udara di Macau.

Dari beberapa penelitian sebelumnya, pendekatan *machine learning* telah banyak digunakan dalam analisis kualitas udara dengan membandingkan berbagai algoritma untuk memperoleh model dengan kinerja terbaik (Patel et al., 2025). Sebagian besar penelitian tersebut berfokus pada perbandingan banyak algoritma dalam satu penelitian (Rajagukguk et al., 2025). Meskipun demikian, penelitian yang secara khusus menekankan pada evaluasi kinerja dua algoritma yang sering digunakan dalam analisis kualitas udara, yaitu *Random Forest* dan *XGBoost*, masih relatif terbatas. Selain itu, sebagian besar penelitian sebelumnya dilakukan pada wilayah di luar Indonesia, sehingga karakteristik data yang digunakan berbeda dengan kondisi kualitas udara di Kota Bandar

Lampung. *Random Forest* dan *XGBoost* dipilih dalam penelitian ini karena pada berbagai penelitian sebelumnya kedua algoritma tersebut menunjukkan performa yang baik dalam menangani data kualitas udara yang kompleks dan bersifat fluktuatif.

Berdasarkan kondisi tersebut, penelitian ini difokuskan pada evaluasi kinerja algoritma *Random Forest* dan *XGBoost* dalam menganalisis kualitas udara di Kota Bandar Lampung. Penelitian ini menggunakan data periode 2023–2025 yang diperoleh dari stasiun pemantauan AQMS yang dikelola oleh Dinas Lingkungan Hidup dan berlokasi di wilayah Sukarame. Melalui perbandingan kinerja kedua algoritma tersebut, penelitian ini diharapkan dapat memberikan gambaran mengenai algoritma yang paling sesuai digunakan dalam analisis kualitas udara berdasarkan data pemantauan di Kota Bandar Lampung serta memberikan informasi yang dapat mendukung upaya pemantauan dan pengelolaan kualitas udara di Kota Bandar Lampung, khususnya pada wilayah Sukarame dan sekitarnya.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang, rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana hasil analisis kualitas udara di Kota Bandar Lampung menggunakan algoritma *Random Forest* dan *XGBoost*?
2. Bagaimana perbandingan kinerja algoritma *Random Forest* dan *XGBoost* dalam menganalisis kualitas udara di Kota Bandar Lampung?

1.3 Batasan Masalah

Adapun batasan masalah penelitian ini adalah sebagai berikut.

1. Pendekatan *machine learning* yang digunakan dalam penelitian ini adalah algoritma *Random Forest* dan *XGBoost*.
2. Data yang digunakan merupakan data kualitas udara dari stasiun pemantauan AQMS di wilayah Sukarame yang dikelola oleh Dinas Lingkungan Hidup Kota Bandar Lampung.
3. Data yang digunakan mencakup periode 2023-2025.

4. Variabel yang digunakan meliputi parameter kualitas udara yaitu PM_{10} , $PM_{2,5}$, SO_2 , CO , O_3 , dan NO_2 , serta parameter meteorologi yaitu kelembaban, kecepatan angin, dan suhu.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut.

1. Mengevaluasi kinerja algoritma *Random Forest* dan *XGBoost* untuk menganalisis kualitas udara Kota Bandar Lampung.
2. Membandingkan performa kinerja algoritma *Random Forest* dan *XGBoost* dalam menganalisis kualitas udara di Kota Bandar Lampung.

1.5 Manfaat Penelitian

Manfaat yang akan diperoleh dari penelitian ini antara lain sebagai berikut.

1. Memberikan informasi mengenai hasil analisis kualitas udara di Kota Bandar Lampung untuk mendukung upaya pemantauan dan pengendalian pencemaran udara.
2. Memberikan referensi mengenai kinerja algoritma *Random Forest* dan *XGBoost* dalam analisis kualitas udara yang dapat digunakan oleh peneliti selanjutnya.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Terdapat beberapa penelitian terdahulu yang berhubungan dengan penelitian ini yang diuraikan dalam Tabel 1.

Tabel 1. Penelitian Terdahulu

No	Penelitian	Objek	Metode	Hasil
1	<i>Air quality prediction by machine learning models: A predictive study on the Indian coastal city of Visakhapatnam</i> (Ravindiran et al., 2023).	Data kualitas udara Kota Visakhapatnam, India (2017–2022).	<i>LightGBM, Random Forest, CatBoost, AdaBoost, dan XGBoost.</i>	<i>CatBoost</i> terbaik ($R^2 = 0,9998$; RMSE = 0,76). <i>Random Forest</i> juga sangat baik ($R^2 = 0,9990$). <i>AdaBoost</i> terendah ($R^2 = 0,9753$).
2	<i>Machine learning-based forecasting of air quality index under long-term environmental patterns</i> (Tırink, 2025).	Data kualitas udara Kota Iğdır, Türkiye (2016–2024).	<i>XGBoost, LightGBM, SVM.</i>	<i>XGBoost</i> terbaik ($R^2 = 0,994$; RMSE = 4,84). <i>LightGBM</i> cukup baik ($R^2 = 0,889$). <i>SVM</i> terendah ($R^2 = 0,782$).
3	<i>Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction</i> (Özüpak et al., 2025).	Dataset <i>Kaggle Air Quality Italia</i> (2004–2005).	<i>XGBoost, LightGBM, Random Forest, Gradient Boosting, CatBoost, SVR.</i>	<i>SVR (Bayesian Optimized)</i> terbaik ($R^2 = 0,9994$; RMSE = 0,022). <i>Gradient Boosting & Random Forest</i> juga sangat tinggi ($>0,998$).

No	Penelitian	Objek	Metode	Hasil
4	<i>Air Pollution Forecasting in Almaty using Ensemble Machine Learning Models</i> (Naizabayeva et al., 2025).	Data kualitas udara Almaty, Kazakhstan (2020–2023).	<i>XGBoost</i> , <i>LightGBM</i> , <i>Weighted Ensemble</i> (XGB + LGBM).	<i>Weighted Ensemble</i> terbaik (R^2 PM _{2,5} = 0,9825; MAE = 0,001). <i>XGBoost</i> kedua (R^2 = 0,9762). <i>LightGBM</i> terendah (R^2 = 0,9676).
5	<i>Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau</i> (Lei et al., 2023).	Data polutan & meteorologi Macau (2016–2021).	ANN, <i>Random Forest</i> , <i>XGBoost</i> , SVM, MLR.	SVM terbaik untuk PM ₁₀ & CO (R^2 = 0,94). Untuk PM _{2,5} , <i>Random Forest</i> paling stabil (R^2 = 0,88). <i>XGBoost</i> sangat baik (0,90).
6	Analisis dan Prediksi Indeks Kualitas Udara Jakarta dengan <i>XGBoost</i> (Sari et al., 2025).	Data AQI harian Jakarta (2024–2025).	<i>XGBoost</i> .	<i>Training</i> sangat baik (R^2 = 0,9917). <i>Testing</i> tetap tinggi (R^2 = 0,9831; RMSE = 2,94). Tidak <i>overfitting</i> .
7	<i>HExplainable Forecasting of Air Quality Index Using a Hybrid Random Forest and ARIMA Model</i> (Yenkikar et al., 2025).	Data AQI India (2015–2020).	<i>Random Forest</i> , ARIMA, Hybrid RF+ARIMA	Hybrid RF+ARIMA terbaik (R^2 = 0,940). <i>Random Forest</i> tunggal lebih rendah (0,917). Faktor paling berpengaruh: PM _{2,5} , PM ₁₀ , NO ₂ .
8	<i>Machine Learning Techniques to Predict the Air Quality Using Meteorological Data in Two Urban Areas in Sri Lanka</i> (Mampitiya et al., 2023).	Data Kandy & Battaramulla (2019–2021).	<i>CatBoost</i> , <i>LightGBM</i> , <i>XGBoost</i> , LSTM, GRU.	<i>CatBoost</i> akurasi tertinggi (Kandy R^2 = 0,99997; Battaramulla R^2 = 0,9999). <i>LightGBM</i> stabil. <i>XGBoost</i> /GRU/LS TM lebih rendah.

No	Penelitian	Objek	Metode	Hasil
9	<i>S-GBR Model Incorporating Seasonal Variables</i> (Mahbodi & Karasfi, 2025).	Data polutan dan meteorologi (fitur musiman).	RF, <i>Gradient Boosting</i> , <i>XGBoost</i> , <i>LightGBM</i> , <i>CatBoost</i> , S-GBR.	S-GBR terbaik (R^2 musim 0,9559–0,9686). <i>XGBoost</i> meningkat signifikan setelah optimasi. RF stabil tetapi tidak tertinggi.
10	<i>Air Quality Forecasting: A Global Perspective with Relevance to Low-Resource Settings</i> (Christian & Choi, 2024).	Dataset global 197 ibu kota (2023).	<i>Linear Regression</i> , <i>Ridge</i> , SVR, <i>Decision Tree</i> , <i>Random Forest</i> , <i>XGBoost</i> , <i>LightGBM</i> , <i>Logistic Regression</i> , <i>KNN</i> , <i>SVC</i> .	<i>Random Forest</i> terbaik untuk regresi ($R^2 = 0,91$); <i>XGBoost</i> baik (0,88). Pada klasifikasi, <i>Random Forest</i> akurasi 1,00; <i>XGBoost</i> 0,99; <i>Decision Tree</i> 1,00 namun dilaporkan mengalami <i>overfitting</i> .

Model prediksi kualitas udara terus berkembang seiring meningkatnya kebutuhan akan sistem pemantauan lingkungan yang lebih akurat. Dalam konteks ini, berbagai studi telah mengevaluasi efektivitas algoritma *machine learning* dalam memprediksi perubahan polutan. Salah satunya terlihat pada penelitian (Ravindiran et al., 2023) yang mengkaji kualitas udara di Kota Visakhapatnam, India, menggunakan data tahun 2017–2022. Dengan membandingkan *LightGBM*, *Random Forest*, *CatBoost*, *AdaBoost*, dan *XGBoost*, penelitian tersebut menunjukkan bahwa *CatBoost* menjadi model paling unggul berdasarkan nilai R^2 sebesar 0,9998 dan RMSE 0,76, sedangkan *AdaBoost* menjadi model yang memiliki akurasi terendah. Hasil ini menegaskan bahwa pendekatan *gradient boosting*, khususnya *CatBoost*, mampu menangkap variasi polutan dengan sangat baik.

Kinerja metode *boosting* yang kuat juga terlihat pada penelitian (Tırınk, 2025) yang memprediksi *Air Quality Index* (AQI) di Kota Iğdır, Türkiye, menggunakan data polutan dan meteorologi selama periode 2016–2024. Dengan

membandingkan *XGBoost*, *LightGBM*, dan *SVM*. Hasil evaluasi menunjukkan bahwa *XGBoost* menjadi model dengan performa tertinggi, memperoleh nilai R^2 0,994 dan RMSE 4,84. Penelitian ini menegaskan keunggulan *XGBoost* dalam menangkap pola dan perubahan polutan secara lebih efektif dibandingkan dua model lainnya.

Hasil yang sejalan juga ditunjukkan oleh (Özüpak et al., 2025) melalui pemanfaatan dataset kualitas udara Italia tahun 2004–2005. Penelitian ini menguji sepuluh algoritma, termasuk *XGBoost*, *LightGBM*, *Random Forest*, *Gradient Boosting*, *CatBoost*, *SVR*, *Decision Tree*, *KNN*, *Elastic Net*, dan *Bayesian Ridge*. Seluruh model dioptimalkan menggunakan *Bayesian Optimization* dan divalidasi menggunakan *Randomized Cross-Validation*. Dari pengujian tersebut, enam algoritma *XGBoost*, *LightGBM*, *Random Forest*, *CatBoost*, *Gradient Boosting*, dan (*Support Vector Regression*) *SVR*. Hasil menunjukkan performa yang unggul, dengan (*Support Vector Regression*) *SVR* berada di posisi teratas melalui nilai R^2 0,9994 dan RMSE 0,022. Hal tersebut memperkuat *tuning hyperparameter* dalam meningkatkan akurasi model.

Kemampuan metode *boosting* di perkuat dalam penelitian (Naizabayeva et al., 2025) di Kota Almaty, Kazakhstan. Data polutan dan meteorologi yang diperoleh dari 20 stasiun pemantauan diolah menggunakan metode *STL (Seasonal-Trend decomposition using Loess)* untuk memisahkan komponen musiman sebelum proses pemodelan. Penelitian ini membandingkan *XGBoost*, *LightGBM*, dan model *ensemble* yang dibuat melalui penggabungan *XGBoost* dan *LightGBM*. Hasil terbaik diperoleh dari model *ensemble* (gabungan *XGBoost* dan *LightGBM*) dengan nilai R^2 0,9825 untuk $PM_{2,5}$ dan 0,8391 untuk CO, Penelitian ini menguatkan bahwa kombinasi *XGBoost* dan *LightGBM* mampu memberikan performa prediksi yang sangat akurat pada data kualitas udara yang memiliki pola musiman yang kuat.

Hasil yang akurat di dukung oleh penelitian (Lei et al., 2023) di Macau yang memprediksi konsentrasi $PM_{2,5}$, PM_{10} , dan CO dengan menggunakan *Artificial Neural Network (ANN)*, *Random Forest*, *XGBoost*, *Support Vector Machine (SVM)*, dan *Multiple Linear Regression (MLR)*. Dengan memanfaatkan data

tahun 2016–2021, SVM menjadi model paling unggul untuk PM_{10} dan CO, masing-masing dengan nilai R^2 0,94. Sementara itu, *Random Forest* memberikan performa paling stabil untuk $PM_{2,5}$ dengan nilai R^2 0,88. Penelitian ini mengindikasikan bahwa setiap polutan memiliki karakteristik yang berbeda sehingga memerlukan model prediksi yang sesuai.

Tingkat performa tinggi juga terlihat pada penelitian (Sari et al., 2025) yang memprediksi Indeks Kualitas Udara (AQI) di Jakarta menggunakan model *XGBoost* dengan dataset AQI harian tahun 2024–2025. Model yang dikembangkan menghasilkan performa yang sangat baik dengan nilai R^2 0,9917 pada data pelatihan dan 0,9831 pada data pengujian. Nilai tersebut menunjukkan bahwa *XGBoost* mampu mempelajari pola perubahan AQI secara stabil tanpa mengalami *overfitting*.

Upaya untuk meningkatkan akurasi juga dilakukan melalui pendekatan *hybrid* dalam penelitian (Yenkikar et al., 2025). Kombinasi *Random Forest* dan ARIMA (*Autoregressive Integrated Moving Average*) digunakan untuk memprediksi AQI di India, di mana *Random Forest* berperan dalam mempelajari hubungan nonlinear antarpolutan, sementara ARIMA menangani residu temporal. Model *hybrid* ini mencatat nilai R^2 0,94, lebih tinggi dibandingkan *Random Forest* tunggal (0,917). Hasil analisis SHAP mengidentifikasi $PM_{2,5}$, PM_{10} , dan NO_2 sebagai variabel paling berpengaruh.

Konsistensi performa algoritma *boosting* turut diperlihatkan dalam penelitian (Mampitiya et al., 2023) yang memprediksi PM_{10} di dua wilayah urban Sri Lanka, yaitu Kandy dan Battaramulla, menggunakan *CatBoost*, *LightGBM*, *XGBoost*, *LSTM*, dan *GRU*. Dari seluruh model yang diuji, *CatBoost* memperoleh akurasi paling tinggi, dengan R^2 sebesar 0,99997 di Kandy dan 0,9999 di Battaramulla. *LightGBM* menunjukkan performa yang stabil pada kedua wilayah, sementara *XGBoost*, *LSTM*, dan *GRU* memberikan hasil yang lebih rendah.

Pengembangan model *Seasonal Gradient Boosting Regressor* (S-GBR) dilakukan dalam penelitian (Mahbodi & Karasfi, 2025) dengan memanfaatkan

penambahan fitur musiman serta penerapan *Bayesian Optimization*. Hasil penelitian menunjukkan bahwa S-GBR mampu memberikan performa terbaik dibandingkan model lainnya dengan nilai R^2 0,9559–0,9686. Hal ini menunjukkan bahwa penambahan fitur musiman dapat meningkatkan ketepatan prediksi kualitas udara.

Penerapan algoritma *ensemble* dalam skala yang lebih luas ditunjukkan oleh (Christian & Choi, 2024) melalui analisis kualitas udara global berbasis data dari 197 ibu kota dunia. Penelitian tersebut menggunakan berbagai algoritma, termasuk *Linear Regression*, *Ridge Regression*, *SVR*, *Decision Tree*, *Random Forest*, *XGBoost*, *LightGBM*, *Logistic Regression*, *KNN*, dan *SVC*. Pada pemodelan regresi, *Random Forest* menjadi model paling unggul dengan nilai R^2 sebesar 0,91, diikuti oleh *XGBoost* yang juga menunjukkan performa baik dengan R^2 sebesar 0,88. Untuk tugas klasifikasi, *Random Forest* kembali memberikan performa terbaik dengan akurasi mencapai 1,00, sedangkan *XGBoost* memperoleh akurasi sangat tinggi, yaitu 0,99. Temuan ini menegaskan bahwa algoritma *ensemble* berbasis pohon memiliki kemampuan prediktif yang kuat ketika diterapkan pada dataset multikota dengan karakteristik polusi yang beragam.

2.2 Landasan Teori

2.2.1 Kota Bandar Lampung

Kota Bandar Lampung merupakan ibu kota Provinsi Lampung yang terletak di bagian selatan Pulau Sumatra. Letaknya yang strategis menjadikan kota ini sebagai pintu gerbang utama aktivitas transportasi darat yang menghubungkan Pulau Jawa dan Sumatra. Kedudukan ini menjadikan Bandar Lampung sebagai pusat pemerintahan, perdagangan, jasa, dan logistik dengan tingkat mobilitas penduduk yang tinggi setiap harinya. Beragam aktivitas tersebut semakin meningkat seiring dengan berkembangnya wilayah kota, sehingga memunculkan tantangan baru terkait pengelolaan lingkungan, termasuk dalam aspek kualitas udara (Iqbal et al., 2019). Kota Bandar Lampung memiliki kondisi wilayah yang

beragam dengan karakteristik tertentu. Gambaran umum wilayah Kota Bandar Lampung disajikan pada Gambar 1.



Gambar 1. Kota Bandar Lampung
Sumber : (Pikiran Rakyat, 2024)

Berdasarkan Gambar 1, sebagai salah satu kota besar di Sumatra, Bandar Lampung memiliki jumlah penduduk yang besar dan tingkat kepadatan yang tinggi, dengan aktivitas masyarakat yang tersebar di kecamatan seperti Tanjung Karang Pusat, Kedaton, Sukarame, Panjang, dan Rajabasa. Kondisi ini mendorong meningkatnya kebutuhan transportasi, konsumsi energi, dan berbagai kegiatan ekonomi yang berpotensi menambah beban pencemaran udara. Tingginya intensitas kendaraan bermotor, terutama di kawasan pusat kota, menyebabkan polutan seperti CO, NO₂, dan partikulat mudah menumpuk di udara, khususnya ketika intensitas kendaraan meningkat dan arus lalu lintas menjadi padat (Kiswandono et al., 2025).

Dengan meningkatnya aktivitas mobilitas di Kota Bandar Lampung, pertumbuhan ekonomi dan perkembangan teknologi turut mendorong kebutuhan masyarakat terhadap sarana transportasi yang lebih cepat dan efisien. Hal ini terlihat dari jumlah kendaraan bermotor di Provinsi Lampung yang pada tahun 2022 mendekati 3,9 juta unit, dengan dominasi

kendaraan roda dua. Kondisi ini menunjukkan tingginya ketergantungan masyarakat terhadap transportasi pribadi (Komarudin et al., 2023). Namun, peningkatan jumlah kendaraan ini juga menimbulkan konsekuensi berupa naiknya emisi gas buang yang memberikan tekanan terhadap kualitas udara kota.

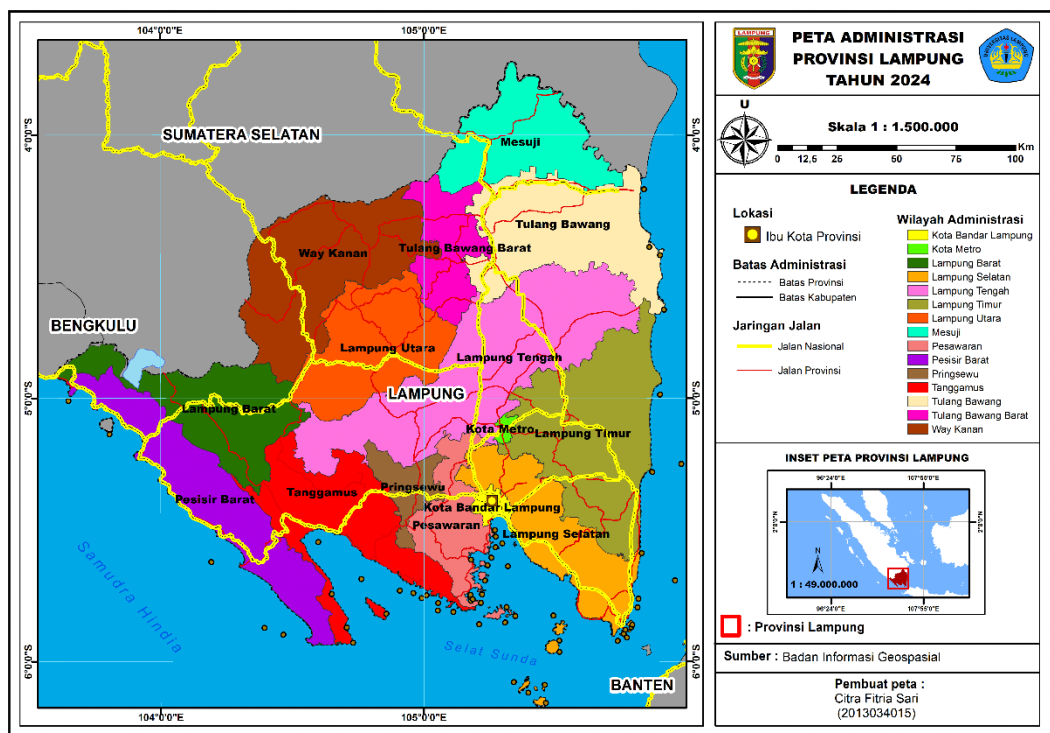
Permasalahan kualitas udara semakin terlihat pada ruas jalan utama seperti Jalan Soekarno Hatta, Yos Sudarso, dan Z.A. Pagar Alam. Jalan-jalan ini memiliki volume lalu lintas yang tinggi dan sering dilalui kendaraan berbobot besar seperti truk dan angkutan logistik. Pergerakan kendaraan berat tersebut dapat meningkatkan konsentrasi debu jalanan dan partikulat (PM_{10} dan $PM_{2,5}$) akibat gesekan roda serta material permukaan jalan yang terangkat ke udara. Kondisi ini menyebabkan kawasan tersebut menjadi titik-titik yang rawan mengalami peningkatan polutan udara

Secara umum, kondisi kualitas udara di berbagai wilayah Kota Bandar Lampung menunjukkan bahwa partikulat menjadi jenis polutan yang paling dominan. Pada sejumlah titik pemantauan kualitas udara di Bandar Lampung, konsentrasi partikulat, khususnya PM_{10} , sering menjadi kontributor utama pada area dengan aktivitas transportasi tinggi. Sementara itu, $PM_{2,5}$ dan polutan gas seperti SO_2 , NO_2 , dan O_3 umumnya berada dalam kisaran yang lebih stabil. Dominasi partikulat ini berkaitan dengan tingginya mobilitas kendaraan, aktivitas permukiman padat, serta debu jalanan yang mudah tersebar oleh pergerakan kendaraan. Polutan partikulat tersebut juga dapat menimbulkan ketidaknyamanan dan memicu iritasi pernapasan bagi masyarakat (Kiswando et al., 2025).

Melihat kondisi tersebut, dapat disimpulkan bahwa Kota Bandar Lampung menghadapi tantangan lingkungan yang cukup kompleks akibat tingginya aktivitas penduduk dan mobilitas transportasi (Iqbal et al., 2019). Peningkatan konsentrasi polutan, khususnya partikulat dari aktivitas lalu lintas dan permukiman padat, menunjukkan perlunya pengelolaan lingkungan yang lebih optimal. Upaya seperti peningkatan ruang terbuka

hijau, penataan transportasi, serta pengendalian sumber emisi menjadi langkah penting untuk menjaga kualitas udara agar lingkungan kota tetap sehat, aman, dan layak huni bagi masyarakat.

Berdasarkan kondisi tersebut, Kota Bandar Lampung tidak terlepas dari wilayah administratif yang lebih luas, yaitu Provinsi Lampung. Oleh karena itu, untuk melihat posisi Kota Bandar Lampung dalam lingkup wilayah yang lebih luas, diperlukan gambaran wilayah Provinsi Lampung. Peta wilayah Provinsi Lampung disajikan pada Gambar 2.

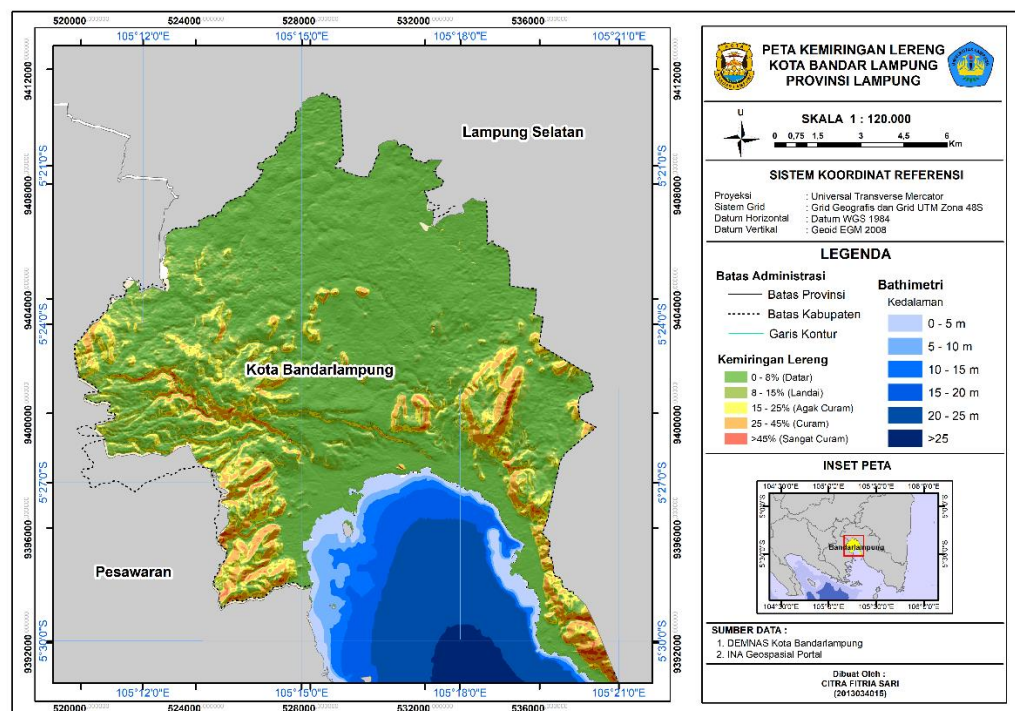


Gambar 2. Peta Provinsi Lampung
Sumber : (Citra Fitria Sari, 2024)

Provinsi Lampung terletak di bagian selatan Pulau Sumatra dan berbatasan langsung dengan Provinsi Sumatera Selatan di sebelah utara serta Selat Sunda di bagian selatan. Berdasarkan peta pada Gambar 2, Provinsi Lampung terdiri atas 13 kabupaten dan 2 kota, dengan Kota Bandar Lampung sebagai ibu kota provinsi.

Pada peta tersebut terlihat bahwa Kota Bandar Lampung berada di bagian selatan Provinsi Lampung dan memiliki posisi yang strategis. Letak ini menjadikan Kota Bandar Lampung sebagai pusat aktivitas pemerintahan, ekonomi, dan transportasi di wilayah Lampung, yang juga berpengaruh terhadap tingginya mobilitas dan potensi pencemaran udara.

Setelah mengetahui posisi Kota Bandar Lampung dalam wilayah Provinsi Lampung, selanjutnya perlu dilihat gambaran wilayah Kota Bandar Lampung secara lebih rinci. Peta wilayah Kota Bandar Lampung disajikan pada Gambar 3.



Gambar 3. Peta Kota Bandar Lampung
Sumber : (Citra Fitria Sari, 2024)

Berdasarkan peta pada Gambar 3, wilayah Kota Bandar Lampung memiliki kondisi topografi yang beragam, mulai dari daerah dataran hingga perbukitan. Beberapa bagian wilayah menunjukkan kemiringan lereng yang cukup curam, terutama pada area tertentu, sementara bagian lainnya merupakan daerah yang relatif landai.

Variasi kondisi topografi ini dapat mempengaruhi aktivitas masyarakat serta pergerakan udara di wilayah tersebut. Daerah dengan kemiringan tertentu berpotensi mempengaruhi penyebaran polutan udara, sehingga kondisi geografis menjadi salah satu faktor yang perlu diperhatikan dalam analisis kualitas udara di Kota Bandar Lampung.

Untuk mendukung pemantauan kualitas udara secara lebih spesifik, diperlukan titik pengamatan yang mampu merepresentasikan kondisi lingkungan di Kota Bandar Lampung. Dalam penelitian ini, pemantauan kualitas udara difokuskan pada stasiun *Automatic Air Quality Monitoring System* (AQMS) yang berlokasi di Kecamatan Sukarame dan dikelola oleh Dinas Lingkungan Hidup Kota Bandar Lampung. Lokasi titik pemantauan kualitas udara tersebut disajikan pada Gambar 4.



Gambar 4. Peta Titik Lokasi Stasiun Pemantauan Kualitas Udara (AQMS) di Kota Bandar Lampung
Sumber : (Wikimedia Commons, 2025)

Berdasarkan peta pada Gambar 4, stasiun AQMS tersebut merupakan satu-satunya titik pemantauan kualitas udara di Kota Bandar Lampung. Data yang dihasilkan dari stasiun ini memiliki jangkauan pemantauan sekitar ± 5 km yang dipengaruhi oleh kondisi meteorologi seperti arah dan kecepatan angin, sehingga mampu merepresentasikan kondisi kualitas udara tidak hanya pada titik lokasi alat, tetapi juga wilayah di sekitarnya. Dengan demikian, titik lokasi stasiun AQMS ini menjadi acuan utama dalam penelitian untuk menggambarkan kondisi kualitas udara di Kota Bandar Lampung.

2.2.2 Kualitas Udara

Kualitas udara merupakan salah satu indikator penting dalam menilai kondisi lingkungan di suatu wilayah, yang mencerminkan tingkat kebersihan udara dari berbagai zat pencemar yang dapat membahayakan kesehatan makhluk hidup dan keseimbangan ekosistem. Menurut *World Health Organization* (WHO), kualitas udara yang baik adalah kondisi di mana konsentrasi polutan di atmosfer berada pada tingkat yang aman dan tidak menimbulkan dampak negatif terhadap kesehatan manusia maupun lingkungan. Penilaian kualitas udara dilakukan dengan mengukur konsentrasi berbagai parameter polutan seperti partikulat (PM_{10} dan $PM_{2,5}$), gas-gas berbahaya (SO_2 , CO, O_3 , dan NO_2), serta kondisi meteorologi yang mempengaruhi penyebaran polutan seperti kelembaban, kecepatan angin, dan suhu (Setiawan et al., 2024).

Udara yang berkualitas baik memiliki komposisi yang seimbang dan tidak mengandung zat berbahaya melebihi ambang batas yang ditetapkan oleh standar nasional maupun internasional. Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup, baku mutu udara ambien merupakan batas kadar zat, energi, atau komponen tertentu yang terdapat di udara ambien, termasuk unsur pencemar yang masih dapat ditoleransi keberadaannya. Penetapan baku mutu ini bertujuan untuk melindungi kesehatan masyarakat, makhluk

hidup lainnya, serta menjaga kelestarian fungsi lingkungan hidup (Aminudin et al., 2021).

Penurunan kualitas udara atau pencemaran udara terjadi ketika konsentrasi satu atau lebih polutan melebihi batas yang aman, sehingga udara tidak lagi layak untuk dihirup atau mendukung kehidupan. Pencemaran udara dapat bersumber dari aktivitas antropogenik (buatan manusia) maupun sumber alami. Aktivitas antropogenik yang menjadi penyumbang utama pencemaran udara antara lain emisi kendaraan bermotor, kegiatan industri, pembakaran bahan bakar fosil di pembangkit listrik, pembakaran sampah, serta aktivitas konstruksi. Sementara itu, sumber alami pencemaran udara dapat berasal dari letusan gunung berapi, kebakaran hutan, debu gurun, serta proses biologis dari tumbuhan dan hewan (Bhathal & Gupta, 2021).

Dampak dari penurunan kualitas udara sangat luas. Dari aspek kesehatan, paparan polutan udara dalam waktu singkat dapat menyebabkan iritasi mata, hidung, dan tenggorokan, serta memicu serangan asma pada individu yang rentan. Sementara itu, paparan dalam waktu lama dapat menimbulkan penyakit yang lebih serius seperti infeksi saluran pernapasan akut (ISPA), bronkitis kronis, emfisema, penyakit jantung koroner, stroke, hingga kanker paru-paru. Menurut data WHO, pencemaran udara menyebabkan sekitar 7 juta kematian prematur setiap tahunnya di seluruh dunia, dengan mayoritas terjadi di negara-negara berkembang yang memiliki tingkat pencemaran tinggi namun infrastruktur kesehatan yang terbatas (Natarajan et al., 2024).

Selain dampak kesehatan, pencemaran udara juga menimbulkan efek negatif terhadap lingkungan. Polutan seperti SO_2 dan NO_2 dapat bereaksi dengan uap air di atmosfer membentuk asam sulfat dan asam nitrat yang kemudian turun sebagai hujan asam. Hujan asam dapat merusak vegetasi, mengubah pH tanah dan perairan, serta mempercepat korosi pada bangunan dan infrastruktur. Selain itu, akumulasi polutan di atmosfer

juga berkontribusi terhadap fenomena perubahan iklim global melalui efek rumah kaca dan pemanasan global (Bhathal & Gupta, 2021). Kondisi ini menunjukkan bahwa pencemaran udara tidak hanya berdampak pada kesehatan manusia, tetapi juga berpengaruh signifikan terhadap keseimbangan lingkungan secara keseluruhan.

Mengingat kompleksitas dan urgensi permasalahan kualitas udara, maka diperlukan sistem pemantauan yang *continue*, akurat, dan komprehensif untuk mengidentifikasi pola temporal dan spasial pencemaran udara. Dengan adanya data pemantauan yang berkualitas, dapat dilakukan analisis mendalam mengenai faktor-faktor yang mempengaruhi kualitas udara sebagai dasar dalam perumusan kebijakan lingkungan yang berbasis bukti ilmiah dan perlindungan kesehatan masyarakat (Yudono et al., 2025).

2.2.3 Polutan

Polutan udara merupakan zat, gas, atau partikel yang berada di atmosfer dan keberadaannya dapat menurunkan kualitas udara serta menimbulkan dampak negatif bagi kesehatan manusia maupun lingkungan. Polutan dapat berasal dari sumber alami seperti debu vulkanik, asap kebakaran hutan, percikan garam laut, hingga proses biologis dari tumbuhan dan mikroorganisme. Namun, sebagian besar polutan berbahaya berasal dari aktivitas manusia seperti pembakaran bahan bakar fosil, transportasi, kegiatan industri, pembangkit listrik, kegiatan konstruksi, serta pembakaran sampah rumah tangga (Muhammad et al., 2025). Untuk menilai besarnya pengaruh berbagai sumber polutan tersebut, diperlukan pengukuran parameter polutan yang mampu menggambarkan kondisi udara secara lebih terukur.

Dalam penelitian kualitas udara, parameter polutan yang umum digunakan meliputi PM_{10} , $PM_{2,5}$, SO_2 , CO , O_3 , dan NO_2 . Selain itu, faktor meteorologis seperti kelembaban, kecepatan angin dan suhu. Setiap parameter memiliki karakteristik serta dampak kesehatan yang berbeda

sehingga perlu dianalisis secara menyeluruh untuk memberikan gambaran kondisi kualitas udara yang lebih akurat (Manisalidis et al., 2020).

a) *Particulate Matter* $\leq 10 \mu\text{m}$ (PM₁₀)

PM₁₀ adalah partikel debu halus yang berukuran kurang dari atau sama dengan 10 mikrometer. Ukurannya sangat kecil sehingga tidak dapat terlihat langsung oleh mata manusia, namun cukup besar untuk masuk ke saluran pernapasan bagian atas seperti hidung dan tenggorokan. PM₁₀ banyak dihasilkan dari debu jalanan, aktivitas konstruksi, proses industri, pembakaran bahan bakar kendaraan, dan pembakaran sampah (Azizah et al., 2025).

Selain menyebabkan iritasi pada saluran pernapasan, paparan PM₁₀ dalam dapat memicu gangguan kesehatan seperti batuk kronis, alergi, infeksi saluran pernapasan, serta memperburuk gejala asma dan bronkitis. Pada musim kemarau atau kondisi udara kering, konsentrasi PM₁₀ biasanya meningkat karena debu mudah terangkat oleh angin. Polutan ini juga dapat mengurangi jarak pandang, menyebabkan kabut debu, serta menurunkan kualitas udara secara signifikan di daerah perkotaan (Ayus et al., 2023).

b) *Particulate Matter* $\leq 2,5 \mu\text{m}$ (PM_{2,5})

PM_{2,5} adalah partikel debu yang lebih halus dibanding PM₁₀, dengan ukuran maksimal 2,5 mikrometer. Ukurannya yang sangat kecil membuat partikel ini mampu menembus jauh ke dalam paru-paru, bahkan masuk ke aliran darah melalui sistem pernapasan. PM_{2,5} berasal dari pembakaran kendaraan bermotor, asap industri, pembakaran batu bara, asap rokok, serta pembakaran biomassa seperti kayu dan dedaunan (Bai et al., 2024).

PM_{2,5} dikenal sebagai salah satu polutan paling berbahaya karena berkaitan dengan peningkatan risiko penyakit serius seperti penyakit

jantung, stroke, gangguan sistem pernapasan, penurunan fungsi paru-paru, hingga kanker paru-paru (Natarajan et al., 2024). Selain itu, konsentrasi $PM_{2.5}$ yang tinggi dapat menyebabkan fenomena kabut asap (*smog*) yang mengurangi jarak pandang dan mengganggu aktivitas masyarakat. Polutan ini menjadi perhatian utama organisasi kesehatan global karena dampaknya yang berlangsung lama dan sangat berpengaruh terhadap kesehatan manusia.

c) Sulfur Dioksida (SO_2)

Sulfur dioksida (SO_2) adalah gas tidak berwarna namun memiliki bau menyengat dan tajam. Gas ini berasal dari pembakaran bahan bakar fosil seperti batu bara, minyak bumi, solar, serta proses industri dan pembangkit listrik. O_2 merupakan salah satu polutan udara yang mudah dikenali karena memiliki bau menyengat serta dapat menimbulkan efek iritasi. (Yudono et al., 2025).

Ketika terhirup, SO_2 dapat mengiritasi hidung, tenggorokan, dan saluran pernapasan, menyebabkan batuk, sesak napas, serta memperburuk kondisi penderita asma. Paparan SO_2 yang berlangsung lama dapat memicu penyakit paru-paru kronis. Secara lingkungan, SO_2 dapat bereaksi dengan uap air di udara membentuk asam sulfat (H_2SO_4), yang kemudian turun sebagai hujan asam. Hujan asam dapat merusak tanaman, mengubah pH tanah dan air, serta mempercepat korosi pada bangunan dan infrastruktur (Manisalidis et al., 2020). Hal ini menunjukkan bahwa SO_2 merupakan polutan udara yang tidak hanya berdampak pada kesehatan manusia, tetapi juga berpengaruh terhadap kerusakan lingkungan.

d) Karbon Monoksida (CO)

Karbon monoksida (CO) adalah gas beracun yang tidak memiliki warna dan bau, sehingga keberadaannya sulit dideteksi tanpa alat

khusus. Gas ini dihasilkan dari pembakaran tidak sempurna bahan bakar seperti bensin, kayu, batu bara, dan gas alam. CO banyak ditemukan di daerah dengan lalu lintas padat, terutama pada waktu aktivitas kendaraan meningkat, serta pada area dengan ventilasi udara yang buruk (Manisalidis et al., 2020).

CO memiliki daya ikat yang sangat kuat terhadap hemoglobin sehingga dapat menghambat transportasi oksigen ke seluruh tubuh. Hal ini menyebabkan gejala seperti pusing, mual, lelah, sesak napas, hingga kehilangan kesadaran. Pada konsentrasi tinggi, paparan CO dalam waktu singkat dapat menyebabkan keracunan berat bahkan kematian. Karena sifatnya yang sangat berbahaya, pemantauan CO penting dilakukan terutama di pusat kota (Setiawan et al., 2024).

e) Ozon (O₃)

Ozon (O₃) adalah gas yang terbentuk melalui reaksi kimia antara sinar matahari dan polutan seperti nitrogen dioksida (NO₂) serta senyawa organik volatil (VOC). Ozon memiliki dua fungsi berbeda tergantung pada lapisan atmosfer tempatnya berada. Di lapisan stratosfer, ozon berperan penting sebagai pelindung bumi dari radiasi ultraviolet. Namun, di lapisan troposfer (permukaan bumi), ozon menjadi polutan berbahaya yang dapat menyebabkan iritasi mata, hidung, dan tenggorokan (Muhammad et al., 2025).

Ozon permukaan dapat menurunkan fungsi paru-paru, memicu batuk, serta memperparah kondisi penderita asma. Konsentrasi ozon biasanya meningkat pada siang hari terutama saat cuaca panas, karena reaksi fotokimia lebih aktif pada kondisi tersebut. Selain itu, ozon juga dapat merusak tanaman, menghambat pertumbuhan tanaman pertanian, serta berkontribusi pada kerusakan ekosistem jika kadarnya terlalu tinggi (Setiawan et al., 2024).

f) Nitrogen Dioksida (NO₂)

Nitrogen dioksida (NO₂) adalah gas berwarna cokelat kemerahan yang dihasilkan dari emisi kendaraan bermotor, proses pembakaran bahan bakar di pabrik, dan pembangkit listrik berbahan bakar fosil. NO₂ adalah salah satu bagian dari kelompok polutan nitrogen oksida (NO_x), yang berperan besar dalam pencemaran udara (Rahmawati et al., 2024).

Paparan NO₂ dapat mengiritasi saluran pernapasan, menurunkan fungsi paru, dan meningkatkan risiko infeksi saluran napas. Paparan NO₂ dalam waktu lama dapat memperburuk kondisi asma, meningkatkan risiko gangguan kardiovaskular, dan berdampak buruk pada perkembangan saluran pernapasan anak-anak. Selain dampak kesehatan, NO₂ juga berkontribusi pada pembentukan ozon permukaan dan partikel halus sekunder, sehingga memperparah kabut fotokimia di kota-kota besar (Binbusayyis et al., 2024).

g) Kelembaban

Kelembaban menggambarkan jumlah uap air yang terdapat di udara dibandingkan dengan jumlah maksimum yang dapat ditampung udara pada suhu tertentu. Kelembaban merupakan faktor penting dalam penyebaran polutan karena memengaruhi perilaku partikel debu dan reaksi kimia atmosfer (Azizah et al., 2025).

Pada kelembaban tinggi, partikel debu seperti PM₁₀ dan PM_{2,5} cenderung menyerap uap air sehingga ukurannya membesar dan lebih cepat mengendap. Sebaliknya, pada kelembaban rendah, partikel-partikel lebih mudah melayang di udara dan meningkatkan risiko pencemaran. RH juga berpengaruh pada kenyamanan manusia, proses pembentukan ozon, serta aktivitas mikroorganisme di atmosfer (Muhammad et al., 2025).

h) Kecepatan Angin

Kecepatan angin memengaruhi pergerakan dan penyebaran polutan di atmosfer. Angin berperan sebagai media alami yang dapat membawa, mengencerkan, atau memindahkan polutan dari satu wilayah ke wilayah lain. Ketika angin bertiup kencang, polutan akan terbawa dan tersebar ke wilayah lain sehingga konsentrasinya di lokasi tersebut menurun. Sebaliknya, angin yang lemah menyebabkan polutan bertahan di satu tempat dan menumpuk, sehingga meningkatkan pencemaran udara (Ali et al., 2025).

Kecepatan angin juga menentukan pola penyebaran polutan secara horizontal maupun vertikal. Dalam penelitian kualitas udara, parameter ini sangat penting untuk memahami dinamika pencemaran dan menganalisis bagaimana polutan bergerak dalam periode tertentu (Sari et al., 2025).

i) Suhu

Suhu merupakan salah satu faktor utama dalam dinamika atmosfer yang mempengaruhi laju reaksi kimia di udara, pergerakan massa udara, serta pembentukan polutan. Suhu yang tinggi dapat mempercepat pembentukan ozon dan meningkatkan volatilitas senyawa organik yang dapat memperburuk kualitas udara. Pada siang hari ketika suhu meningkat, aktivitas fotokimia biasanya menjadi lebih aktif sehingga konsentrasi ozon ikut meningkat (Yenkikar et al., 2025).

Suhu juga berkaitan dengan fenomena inversi suhu, yaitu kondisi ketika lapisan udara hangat menahan udara dingin di bawahnya sehingga polutan terperangkap dan tidak dapat naik ke lapisan atmosfer yang lebih tinggi. Inversi suhu sering terjadi pada pagi hari atau musim kemarau dan menyebabkan peningkatan konsentrasi polutan di permukaan tanah (Mahbodi & Karasfi, 2025).

2.2.4 *Interpolasi Linear*

Interpolasi linear merupakan metode numerik yang digunakan untuk memperkirakan nilai suatu data yang tidak diketahui berdasarkan dua titik data yang telah diketahui (Kusumawardana, 2021). Metode ini bekerja dengan mengasumsikan bahwa perubahan nilai antar dua titik data bersifat *linear*, sehingga dapat direpresentasikan dalam bentuk garis lurus.

Interpolasi linear banyak digunakan dalam pengolahan data deret waktu (*time series*) karena sederhana, mudah diimplementasikan, dan memiliki tingkat kesalahan yang relatif kecil. Selain itu, metode ini mampu menghasilkan estimasi yang cukup baik tanpa mengubah pola dasar dari data.

Dalam konteks data deret waktu, *interpolasi linear* digunakan untuk mengisi data yang hilang (*missing value*), di mana nilai yang tidak tersedia dapat diperkirakan berdasarkan nilai sebelum dan sesudahnya, sehingga pola temporal tetap terjaga.

Secara matematis, *interpolasi linear* dirumuskan sebagai berikut:

$$f_1 = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0) \quad (1)$$

Sumber : (Kusumawardana, 2021)

Keterangan:

- x = titik atau variabel yang akan dicari nilainya
- x_0 = titik pertama yang diketahui
- x_1 = titik kedua yang diketahui
- $f(x_0)$ = nilai fungsi pada titik x_0
- $f(x_1)$ = nilai fungsi pada titik x_1
- $f(x)$ = nilai hasil *interpolasi*

Dengan menggunakan *interpolasi linear*, nilai yang tidak diketahui dapat diperkirakan secara sederhana dan efisien, serta tetap mempertahankan pola perubahan data sehingga metode ini banyak digunakan dalam pengolahan data deret waktu.

2.2.5 Indeks Standar Pencemar Udara (ISPU)

Indeks Standar Pencemar Udara (ISPU) merupakan suatu nilai yang digunakan untuk menggambarkan kondisi kualitas udara ambien pada suatu lokasi tertentu. ISPU dinyatakan dalam bentuk angka tanpa satuan yang merepresentasikan tingkat pencemaran udara berdasarkan dampaknya terhadap kesehatan manusia, makhluk hidup, serta lingkungan.

ISPU digunakan sebagai indikator untuk memberikan informasi kepada masyarakat mengenai tingkat kualitas udara sehingga dapat menjadi dasar dalam pengambilan keputusan terkait aktivitas di luar ruangan.

Berdasarkan Permen LHK No 14 Tahun 2020, parameter yang digunakan dalam perhitungan ISPU meliputi: PM₁₀, PM_{2,5}, SO₂, CO, O₃, dan NO₂

Perhitungan ISPU dilakukan dengan mengonversi nilai konsentrasi masing-masing parameter menjadi nilai indeks berdasarkan rentang tertentu yang telah ditetapkan.

Secara matematis, perhitungan ISPU dirumuskan sebagai berikut:

$$I = \frac{(I_a - I_b)}{(\chi_a - \chi_b)} \times (\chi_x - \chi_b) + I_b \quad (2)$$

Sumber : (Permen LHK, 2020)

Keterangan:

- I = ISPU terhitung
- I_a = ISPU batas atas
- I_b = ISPU batas bawah
- χ_a = Konsentrasi ambien batas atas ($\mu\text{g}/\text{m}^3$)
- χ_b = Konsentrasi ambien batas bawah ($\mu\text{g}/\text{m}^3$)
- χ_x = Konsentrasi ambien nyata hasil pengukuran ($\mu\text{g}/\text{m}^3$)

Rentang nilai ISPU dikategorikan menjadi beberapa tingkat kualitas udara, yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya.

2.2.6 Artificial Intelligence (AI)

Artificial Intelligence (AI) atau kecerdasan buatan merupakan cabang ilmu komputer yang berfokus pada pembuatan sistem yang mampu melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia. Teknologi ini dirancang untuk meniru proses kognitif seperti kemampuan belajar, penalaran, pengenalan pola, pengambilan keputusan, hingga pemecahan masalah. Pada era modern, AI menjadi fondasi bagi banyak teknologi yang berkembang pesat, terutama karena kemampuannya mengolah data dalam jumlah besar dengan cepat dan efisien. Secara konsep, AI bekerja dengan memanfaatkan algoritma yang dibangun dari data historis untuk menemukan pola, sehingga mampu mengidentifikasi hubungan dalam data dan menghasilkan keluaran (*output*) model yang sesuai (Golec et al., 2025).

Ruang lingkup AI sangat luas dan mencakup berbagai bidang seperti *machine learning*, *deep learning*, *computer vision*, *natural language processing (NLP)*, *robotics*, dan *expert systems*. Setiap cabang AI memiliki metode dan karakteristik sendiri, namun seluruhnya bertujuan untuk menciptakan sistem yang dapat belajar dari pengalaman dan meningkatkan performanya dari waktu ke waktu. Misalnya, NLP memungkinkan komputer memahami bahasa manusia, *computer vision* membantu mesin mengenali objek dari gambar dan video, sementara *deep learning* memanfaatkan jaringan saraf tiruan untuk memproses data kompleks secara mendalam (Sharifani & Amini, 2023).

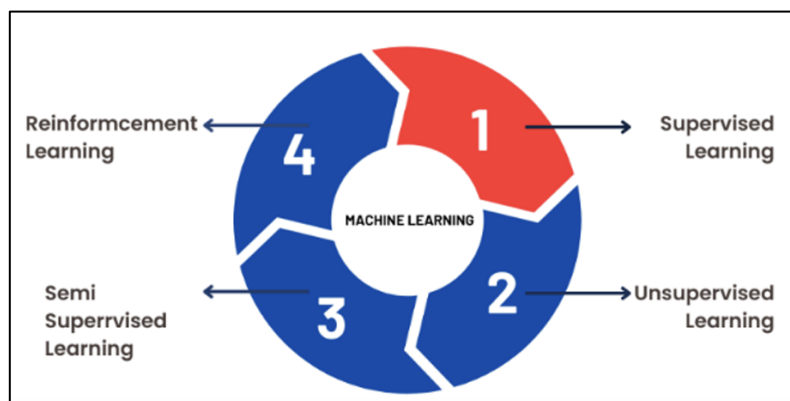
Penggunaan AI telah meluas ke berbagai sektor kehidupan, mulai dari industri manufaktur, perbankan, transportasi, pertanian, pertahanan, hingga kesehatan. Contoh penerapan AI yang sering ditemui adalah asisten virtual seperti *Google Assistant* dan *Siri*, sistem rekomendasi di platform *e-commerce*, deteksi wajah, kendaraan otonom, dan model analisis di berbagai bidang. Dengan perkembangan teknologi *big data* dan komputasi modern, AI menjadi semakin penting dalam

mendukung otomatisasi proses kerja, pengambilan keputusan strategis, serta penyelesaian permasalahan yang rumit secara lebih cepat dan akurat (Soori et al., 2023).

1. *Machine Learning*

Machine Learning (ML) merupakan salah satu cabang utama dari kecerdasan buatan yang memungkinkan komputer belajar dari data tanpa memerlukan instruksi pemrograman secara spesifik. ML bekerja dengan cara mempelajari pola-pola yang terdapat dalam dataset kemudian menggunakannya untuk menghasilkan keputusan atau keluaran (*output*) model terhadap data baru. Kemampuan ML untuk beradaptasi dan meningkatkan performa seiring bertambahnya jumlah data menjadikannya teknologi penting dalam analisis data modern (Razzaq & Shah, 2025).

Berdasarkan cara belajarnya, *machine learning* dibagi menjadi empat kategori utama, disajikan pada Gambar 5.



Gambar 5. Kategori *Machine Learning* Berdasarkan Cara Belajarnya
Sumber : (Cahyani, 2024)

a. *Supervised Learning*

Supervised learning adalah metode pembelajaran mesin yang menggunakan data berlabel, yaitu data yang telah memiliki nilai target atau keluaran (*output*) model. Metode ini bertujuan untuk mempelajari hubungan antara variabel input dan output sehingga model dapat digunakan untuk menganalisis nilai atau kategori pada data baru. Dalam

penerapannya, *supervised learning* umumnya digunakan untuk menyelesaikan dua jenis permasalahan, yaitu klasifikasi (mengelompokkan data ke dalam kategori tertentu) dan regresi (menganalisis nilai numerik). Contoh penerapannya antara lain klasifikasi gambar, penentuan nilai numerik, dan penilaian risiko (Nurhalizah et al., 2024).

b. *Unsupervised Learning*

Unsupervised learning merupakan metode pembelajaran yang menggunakan data tanpa label. Model berusaha menemukan pola, struktur, atau pengelompokan alami dalam data secara mandiri. Contoh penerapannya adalah *clustering*, reduksi dimensi, dan segmentasi data. Metode ini digunakan ketika tidak tersedia informasi target atau ketika ingin memahami struktur dasar data (Retnoningsih & Pramudita, 2020).

c. *Semi-Supervised Learning*

Semi-supervised learning adalah metode pembelajaran yang mengombinasikan sejumlah kecil data berlabel dan sejumlah besar data tanpa label. Pendekatan ini digunakan ketika proses pelabelan sulit, mahal, atau memerlukan keahlian khusus. Dengan memanfaatkan kedua jenis data, *semi-supervised learning* mampu menghasilkan model yang lebih akurat dibandingkan *unsupervised learning*, namun tetap efisien dalam penggunaan data berlabel. Metode ini banyak digunakan pada pemrosesan gambar, teks, dan suara (Shakya et al., 2024).

d. *Reinforcement Learning*

Reinforcement learning adalah metode pembelajaran mesin yang berbasis interaksi antara agen dan lingkungan untuk mencapai tujuan tertentu. Agen belajar melalui sistem penghargaan (*reward*) dan hukuman (*punishment*), sehingga dapat menentukan tindakan terbaik melalui proses percobaan berulang. Tujuan dari metode ini adalah memaksimalkan total reward dalam jangka panjang. Pendekatan ini

banyak digunakan pada bidang robotika, permainan (*game*), dan sistem navigasi (Frommeyer et al., 2025).

2.2.7 *Regression*

Regression atau regresi merupakan metode dalam statistik dan *machine learning* yang digunakan untuk menganalisis hubungan antara dua variabel atau lebih. Metode ini digunakan untuk mengetahui bagaimana suatu variabel memengaruhi variabel lainnya serta memahami pola hubungan yang terdapat dalam data (Akbar et al., 2022).

Tujuan regresi adalah untuk menyusun suatu model yang dapat merepresentasikan hubungan antarvariabel sehingga dapat dijelaskan arah dan kecenderungan hubungan yang terjadi dalam data. Dalam penerapannya, regresi memanfaatkan data yang tersedia untuk menggambarkan hubungan antarvariabel, sehingga dapat diketahui perubahan yang terjadi pada suatu variabel ketika variabel lainnya mengalami perubahan.

Dalam *machine learning*, regresi termasuk ke dalam kategori *supervised learning*, yaitu metode pembelajaran yang menggunakan data berlabel. Hasil dari regresi berupa nilai numerik yang dapat dimanfaatkan dalam berbagai kebutuhan pengolahan dan interpretasi data.

2.2.8 *Random Forest*

Random Forest adalah algoritma *machine learning* berbasis *ensemble learning* yang menggabungkan banyak *decision tree* untuk menghasilkan model yang lebih stabil, akurat, dan tahan terhadap *overfitting*. Teknik *ensemble* ini bekerja dengan membangun sejumlah pohon keputusan secara acak. Setiap pohon dilatih menggunakan *subset* data dan fitur tertentu (metode *bootstrap*), sehingga masing-masing pohon memiliki struktur berbeda. Hasil akhir model ditentukan melalui mekanisme voting pada kasus klasifikasi atau rata-rata pada kasus regresi (Hasnataeni et al., 2025).

Keunggulan *Random Forest* terletak pada kemampuannya menangani data berukuran besar, variabel dalam jumlah banyak, serta hubungan antarvariabel yang kompleks. Algoritma ini juga mampu mengukur tingkat kepentingan setiap fitur (*feature importance*), sehingga dapat membantu peneliti memahami variabel mana yang paling memengaruhi hasil analisis model pengujian. Hal ini sangat bermanfaat dalam penelitian di bidang lingkungan, kesehatan, keuangan, dan berbagai sektor lainnya.

Random Forest memiliki ketahanan yang tinggi terhadap noise dan data yang tidak seimbang, karena proses randomisasi memungkinkan model tetap stabil meskipun terdapat variasi pada data latih. Dengan kombinasi akurasi tinggi, fleksibilitas, dan *interpretabilitas* yang cukup baik, *Random Forest* menjadi salah satu algoritma yang paling populer dan banyak digunakan dalam penelitian dan pengolahan data (Ravindiran et al., 2023).

Untuk memperjelas mekanisme kerja *Random Forest* secara matematis, diperlukan suatu bentuk perumusan yang menggambarkan bagaimana hasil dari setiap pohon keputusan digabungkan. Setiap pohon keputusan dibangun dari sebagian data yang tersedia sehingga menghasilkan keluaran (*output*) model yang bervariasi.

Secara matematis, *Random Forest* dapat dinyatakan sebagai berikut:

$$\text{Random forest prediction} = \frac{1}{K} \sum_{k=1}^K h_k(x) \quad (3)$$

Keterangan :

- K = jumlah total pohon keputusan (*decision tree*)
- $h_k(x)$ = keluaran (*output*) model dari pohon ke- k untuk input x
- $\frac{1}{K} \sum_{k=1}^K h_k(x)$ = rata-rata dari seluruh keluaran (*output*) model yang dihasilkan oleh setiap pohon keputusan

Keluaran (*output*) model yang dimaksud pada rumus tersebut merupakan nilai yang dihasilkan oleh setiap pohon keputusan berdasarkan data input

yang diproses oleh model. Nilai-nilai tersebut kemudian digabungkan dengan cara dirata-ratakan untuk memperoleh hasil akhir *Random Forest*. Pendekatan ini membuat model lebih stabil dan mampu mengurangi risiko *overfitting* dibandingkan dengan penggunaan satu pohon keputusan saja.

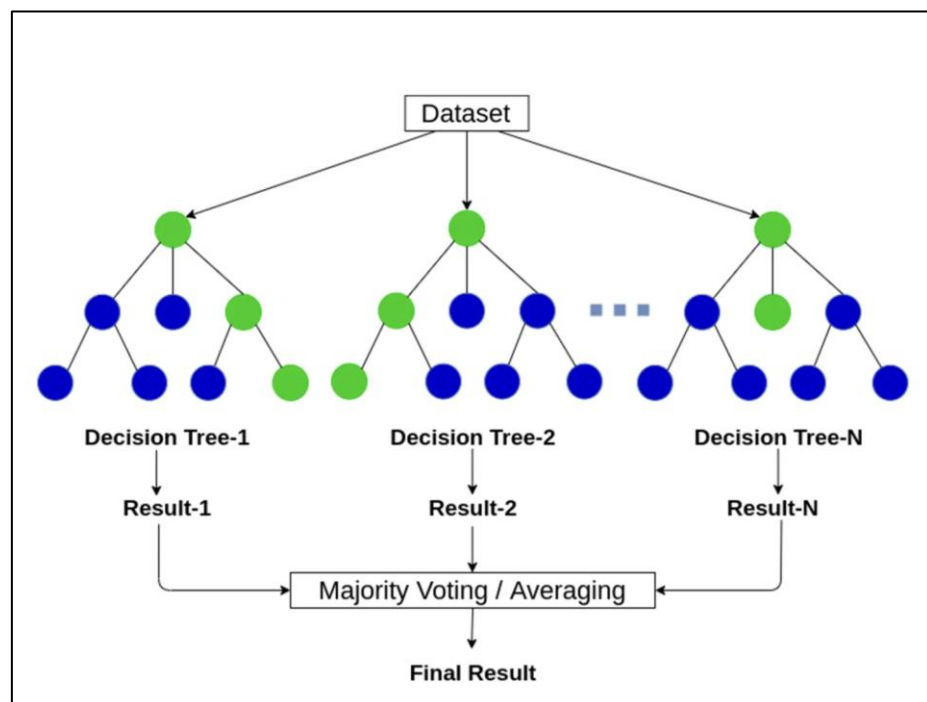
Metode *Random Forest Regression* mampu menghasilkan akurasi yang lebih tinggi dibandingkan *linear regression*. Pada metode ini digunakan rumus khusus dalam proses perhitungannya, yang dinyatakan sebagai berikut :

$$\hat{Y}_i = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{Y}_n \quad (4)$$

Keterangan :

- \hat{Y}_i = nilai keluaran (*output*) model untuk data ke-i,
- N_{tree} = jumlah total pohon keputusan dalam Random Forest,
- \hat{Y}_n = nilai keluaran (*output*) model dari pohon ke-n

Untuk memperjelas alur kerja algoritma *Random Forest*, diagram proses pembentukan model dan penggabungan hasil disajikan pada Gambar 6.



Gambar 6. Alur Algoritma *Random Forest*
Sumber : (Arifianti & Salam, 2024)

Berdasarkan Gambar 6, alur pada diagram menunjukkan bagaimana *Random Forest* bekerja dengan membangun beberapa pohon keputusan secara paralel. Proses dimulai dari satu dataset yang kemudian secara acak dibagi menjadi beberapa bagian melalui teknik *bootstrap sampling*. Setiap bagian data tersebut digunakan untuk melatih satu pohon keputusan, sehingga terbentuk *Decision Tree-1*, *Decision Tree-2*, hingga *Decision Tree-N*. Pada tahap ini, setiap pohon dilatih secara terpisah dan menggunakan sebagian fitur yang dipilih secara acak pada setiap pemecahan node, sehingga masing-masing pohon memiliki struktur yang berbeda.

Setelah seluruh pohon selesai dilatih, masing-masing pohon menghasilkan keluaran (*output*) model sendiri (Result-1 hingga Result-N). Hasil-hasil tersebut kemudian digabungkan melalui proses *majority voting* untuk kasus klasifikasi atau *averaging* untuk kasus regresi. Dengan mengombinasikan berbagai keluaran (*output*) model dari banyak pohon, *Random Forest* mampu menghasilkan hasil akhir yang lebih stabil, *robust*, dan akurat dibandingkan model pohon tunggal. Pendekatan *ensemble* ini juga membuat *Random Forest* menjadi metode yang efektif dalam mengurangi risiko *overfitting* serta meningkatkan kemampuan generalisasi model.

2.2.9 *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) adalah algoritma *ensemble learning* yang merupakan pengembangan dari metode *Gradient Boosting*. Algoritma ini bekerja dengan membangun model secara bertahap, di mana setiap model baru berfungsi untuk memperbaiki kesalahan keluaran (*output*) model sebelumnya. Proses pembelajaran dilakukan secara *additive* dan berfokus pada minimisasi fungsi kerugian (*loss function*).

Keunggulan utama *XGBoost* terletak pada efisiensi komputasi, kemampuan mengelola data berukuran besar, serta ketahanannya terhadap *overfitting* melalui mekanisme regularisasi. Algoritma ini juga

mendukung penanganan *missing value* secara otomatis dan memiliki fleksibilitas dalam mengatur parameter seperti *learning rate*, kedalaman pohon, jumlah pohon, dan fungsi kerugian yang digunakan. Dalam prosesnya, *XGBoost* membangun banyak pohon keputusan (*decision tree*) secara berurutan. Setiap pohon bertugas memperbaiki error dari pohon sebelumnya, sehingga keluaran (*output*) akhir model menjadi lebih akurat. Model akhir diperoleh dengan menjumlahkan kontribusi seluruh pohon yang telah dibangun (Mampitiya et al., 2023). Proses pembentukan keluaran (*output*) model pada setiap iterasi tersebut dapat dijelaskan secara matematis melalui suatu persamaan yang menggambarkan hubungan antara keluaran (*output*) model sebelumnya dan kontribusi pohon baru yang ditambahkan.

Rumus *XGBoost* adalah sebagai berikut:

$$\hat{y}_i^{(t)} = \sum_{k=1}^k f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (5)$$

Keterangan :

- $\hat{y}_i^{(t)}$ = keluaran (*output*) model untuk data ke-i pada iterasi ke-t
- $\hat{y}_i^{(t-1)}$ = keluaran (*output*) model sebelum iterasi ke-t
- $f_t(x_i)$ = pohon keputusan yang dibangun pada iterasi ke-t
- $f_k(x_i)$ = kontribusi pohon keputusan ke-k terhadap keluaran (*output*) model
- t = jumlah iterasi (jumlah pohon yang sudah dibangun)

Rumus ini menunjukkan bahwa keluaran (*output*) model pada iterasi ke-t diperoleh dari keluaran (*output*) model sebelumnya yang ditambah dengan kontribusi pohon baru yang dibentuk pada iterasi tersebut. Melalui proses penambahan secara bertahap ini, *XGBoost* mampu memperbaiki kesalahan pada setiap iterasi.

Model *XGBoost Regression* membentuk keluaran (*output*) model dengan menjumlahkan kontribusi dari setiap pohon keputusan (*regression tree*) yang dibangun selama proses pelatihan (Saleh et al., 2025). Rumus dasar analisis *XGBoost* dituliskan sebagai berikut:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (6)$$

Keterangan :

- \hat{y}_i = nilai keluaran (*output*) model untuk data ke-i
- K = jumlah total pohon keputusan
- $f_k(x_i)$ = keluaran (*output*) model dari pohon ke-k untuk input x_i

Persamaan (6) menunjukkan bahwa keluaran (*output*) akhir model diperoleh dari penjumlahan seluruh keluaran (*output*) model dari pohon yang dibangun.

Pembelajaran *XGBoost* dilakukan secara bertahap (*boosting*). Keluaran (*output*) model pada iterasi ke-t diperbarui dengan menambahkan pohon baru yang dibentuk pada iterasi tersebut. Proses ini menunjukkan bahwa setiap iterasi berkontribusi dalam memperbaiki kesalahan dari keluaran (*output*) model sebelumnya secara bertahap. Secara matematis, proses pembaruan tersebut dapat dinyatakan sebagai berikut:

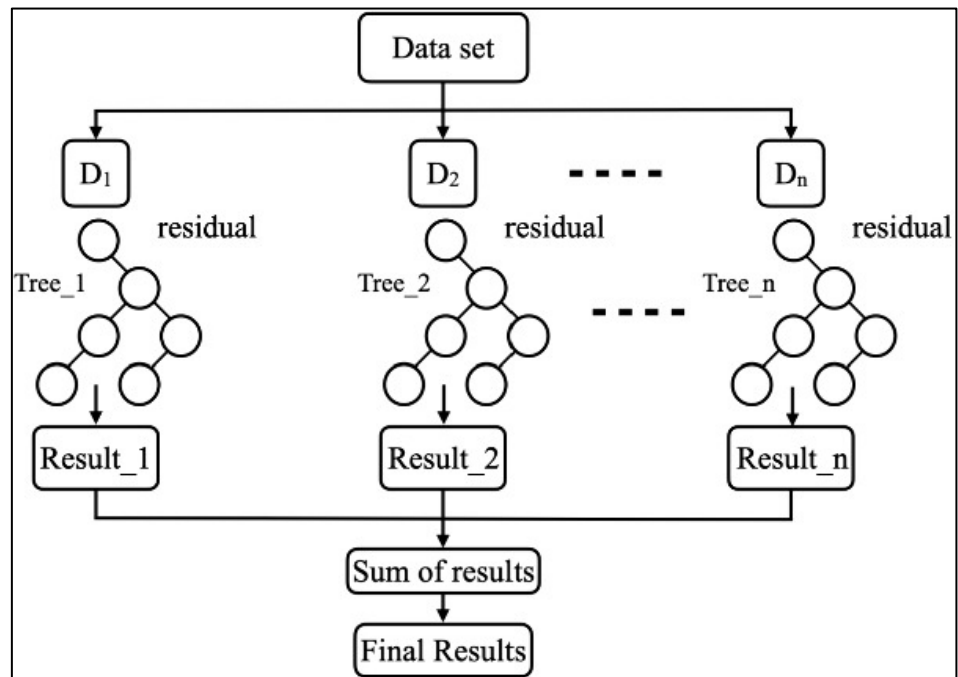
$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (7)$$

Keterangan :

- $\hat{y}_i^{(t)}$ = keluaran (*output*) model pada iterasi ke-t
- $f_t(x_i)$ = pohon keputusan baru yang dibentuk pada iterasi t
- Pohon baru tersebut memperbaiki kesalahan dari keluaran (*output*) model sebelumnya.

Untuk memberikan gambaran yang lebih jelas mengenai mekanisme pembelajaran *XGBoost*, berikut diagram alur pembentukan model pada setiap iterasi yang menggambarkan proses pembelajaran secara bertahap

dalam membangun model melalui pembaruan pada setiap iterasi hingga diperoleh model yang optimal, disajikan pada Gambar 7.



Gambar 7. Alur Algoritma *XGBoost*
Sumber : (Nithin, 2023)

Berdasarkan Gambar 7, alur Alur pada diagram menunjukkan bagaimana *XGBoost* bekerja secara bertahap melalui beberapa iterasi. Proses dimulai dari dataset yang kemudian diproses menjadi D_1 pada iterasi pertama. Pada tahap ini, model menghitung *residual*, yaitu selisih antara nilai asli dalam data dan keluaran (*output*) awal model. Selisih inilah yang digunakan untuk membangun pohon keputusan pertama ($Tree_1$), yang menghasilkan kontribusi keluaran (*output*) model awal ($Result_1$). Setelah itu, *residual* diperbarui berdasarkan hasil keluaran (*output*) model sebelumnya dan digunakan untuk membentuk D_2 pada iterasi berikutnya. Pohon keputusan kedua ($Tree_2$) kemudian dilatih menggunakan *residual* baru tersebut untuk memperbaiki kesalahan dari iterasi sebelumnya sehingga menghasilkan $Result_2$.

Tahapan yang sama terus berlanjut hingga iterasi ke- n , di mana setiap pohon belajar dari sisa kesalahan yang masih ada. Setiap pohon memberikan kontribusi keluaran (*output*) model masing-masing, dan

seluruh hasil tersebut ($Result_1$ hingga $Result_n$) kemudian dijumlahkan untuk menghasilkan keluaran (*output*) akhir model (*Final Results*). Dengan proses perbaikan berulang melalui pembelajaran *residual*, *XGBoost* mampu meningkatkan akurasi model secara bertahap.

2.2.10 Evaluasi Model

Evaluasi model merupakan tahapan penting yang digunakan untuk menilai sejauh mana suatu model mampu menghasilkan keluaran (*output*) model yang mendekati data aktual. Pada pendekatan regresi, proses evaluasi dilakukan dengan mengukur tingkat kesalahan antara nilai aktual dan nilai keluaran (*output*) model.

Untuk menilai kinerja model, digunakan beberapa metrik evaluasi yang umum, yaitu *Mean Absolute Percentage Error* (MAPE), *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Coefficient of Determination* (R^2). Metrik-metrik tersebut digunakan untuk menggambarkan besarnya kesalahan serta kemampuan model dalam merepresentasikan pola data (Chicco et al., 2021). Dengan menggunakan lebih dari satu metrik, evaluasi kinerja model bertujuan untuk memberikan penilaian yang lebih lengkap karena setiap metrik memiliki karakteristik yang berbeda dalam mengukur kesalahan.

a) *Mean Absolute Percentage Error* (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan salah satu ukuran yang digunakan untuk mengetahui tingkat kesalahan suatu model. MAPE menghitung rata-rata persentase perbedaan antara nilai aktual dan nilai keluaran (*output*) model terhadap nilai aktual (Mahbodi & Karasfi, 2025). Dalam perhitungannya, selisih antara nilai aktual dan nilai keluaran (*output*) model diambil tanpa memperhatikan tanda negatif atau positif, sehingga yang diperhatikan hanya besarnya perbedaan. Hal ini bertujuan agar seluruh kesalahan dapat dihitung secara konsisten.

Nilai MAPE yang lebih kecil menunjukkan bahwa keluaran (*output*) model semakin mendekati nilai aktual. Oleh karena itu, MAPE sering digunakan sebagai salah satu indikator dalam mengevaluasi kinerja model karena hasilnya dinyatakan dalam bentuk persentase yang mudah dipahami.

Secara matematis, MAPE dirumuskan sebagai berikut:

$$MAPE = \frac{1}{n} \left(\sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \right) \times 100\% \quad (10)$$

Keterangan :

- n = jumlah data yang digunakan dalam perhitungan.
- Z_t = nilai aktual atau nilai pengamatan pada periode ke- t .
- \hat{Z}_t = nilai keluaran (*output*) model pada periode ke- t .

Kriteria keakuratan dari metode perhitungan kesalahan menggunakan MAPE dapat dilihat pada Tabel 2.

Tabel 2. Kriteria Tingkat Akurasi berdasarkan Nilai MAPE

Nilai MAPE	Kategori Kinerja Model
<10%	Sangat baik
10-20%	Baik
20-50%	Cukup
>50%	Kurang baik

Sumber : (Kurniasari et al., 2024)

b) *Root Mean Square Error (RMSE)*

Root Mean Square Error (RMSE) adalah metrik evaluasi yang digunakan untuk mengukur besarnya kesalahan antara nilai aktual dan nilai keluaran (*output*) model. RMSE diperoleh dengan menghitung akar kuadrat dari rata-rata selisih kuadrat antara kedua nilai tersebut (Tırınk, 2025).

Dalam perhitungannya, selisih nilai dikuadratkan sehingga kesalahan yang besar akan memberikan pengaruh yang lebih tinggi dibandingkan kesalahan yang kecil. Hal ini menyebabkan RMSE lebih sensitif terhadap data yang memiliki penyimpangan besar (*outlier*). Semakin kecil nilai RMSE, semakin dekat keluaran (*output*) model dengan nilai aktual. Oleh karena itu, RMSE sering digunakan sebagai indikator dalam mengevaluasi kinerja model karena mampu menunjukkan besarnya kesalahan secara lebih jelas.

Secara matematis, RMSE dirumuskan sebagai berikut:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (8)$$

Keterangan :

- y_i = nilai aktual pada data ke-i
- \hat{y}_i = nilai keluaran (*output*) model pada data ke-i
- n = jumlah data

c) *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) merupakan metrik evaluasi pada regresi yang digunakan untuk mengetahui rata-rata besarnya kesalahan antara nilai aktual dan nilai keluaran (*output*) model. MAE menghitung selisih antara kedua nilai tersebut tanpa memperhatikan tanda negatif atau positif, sehingga tidak membedakan apakah kesalahan yang terjadi bernilai lebih tinggi atau lebih rendah (Naizabayeva et al., 2025).

Dalam perhitungannya, setiap kesalahan dihitung secara langsung sehingga hasilnya lebih mudah dipahami. Nilai MAE memberikan gambaran rata-rata besarnya kesalahan yang dihasilkan oleh model. Nilai MAE yang lebih kecil menunjukkan bahwa tingkat kesalahan yang keluaran (*output*) model semakin rendah, sehingga hasil yang diperoleh semakin mendekati nilai aktual.

Rumus MAE adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

Keterangan :

- $|y_i - \hat{y}_i|$ = kesalahan absolut
- n = jumlah data

d) Coefficient of Determination (R^2)

Koefisien Determinasi atau R^2 adalah metrik evaluasi yang digunakan untuk mengukur seberapa baik model regresi dapat menjelaskan variabilitas pada data aktual. Nilai R^2 berada pada rentang 0 hingga 1. Nilai mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data, sedangkan nilai mendekati 0 menunjukkan kemampuan model yang rendah dalam memodelkan hubungan antarvariabel (Özüpak et al., 2025).

Rumus R^2 adalah sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

Keterangan :

- y_i = nilai aktual
- \hat{y}_i = nilai keluaran (*output*) model
- \bar{y} = nilai rata-rata dari seluruh data aktual
- $\sum (y_i - \hat{y}_i)^2$ = jumlah kuadrat selisih keluaran model (SSE – error)
- $\sum (y_i - \bar{y})^2$ = jumlah kuadrat selisih terhadap rata-rata (SST – total variasi)

R^2 menunjukkan proporsi variasi dalam data yang mampu dijelaskan oleh model. Nilai mendekati 1 menunjukkan model sangat baik, sedangkan nilai mendekati 0 menunjukkan model kurang mampu menjelaskan data.

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu

3.1.1 Tempat Penelitian

Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang beralamat di Jl. Prof. Dr. Sumantri Brojonegoro No. 1, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung.

3.1.2 Waktu Penelitian

Penelitian ini dilaksanakan pada semester ganjil sampai semester genap dengan perkiraan waktu bulan November 2025 sampai dengan bulan April 2026. Jadwal kegiatan penelitian yang akan dilakukan dapat di sajikan pada Tabel 3.

Tabel 3. Waktu Penelitian

No	Jenis Kegiatan	BULAN PENELITIAN PELAKSANAAN						
		November	Desember	Januari	Februari	Maret	April	Mei
1	Penulisan Laporan Bab 1-3	■						
2	Pengumpulan Data	■						
3	Seminar Usul		■					
4	Preprocessing Data			■				
5	Pembagian Data Training dan Data Testing				■	■	■	■
6	Model <i>Random Forest</i> , dan <i>XGBoost</i>					■	■	■
7	Pengujian					■	■	■

No	Jenis Kegiatan	BULAN PENELITIAN PELAKSANAAN						
		November	Desember	Januari	Februari	Maret	April	Mei
8	Perbandingan Hasil Pengujian Model							
9	Penulisan Laporan Bab 4-5							
10	Seminar Hasil Penelitian							

3.2 Alat Pendukung

3.2.1 Perangkat Keras

- A. Processor: *AMD Ryzen 5 5600H with Radeon Graphics* (6 core, 12 thread)
- B. Memori: 16.0 GB RAM
- C. System Type: 64-bit Operating System, x64-based Processor

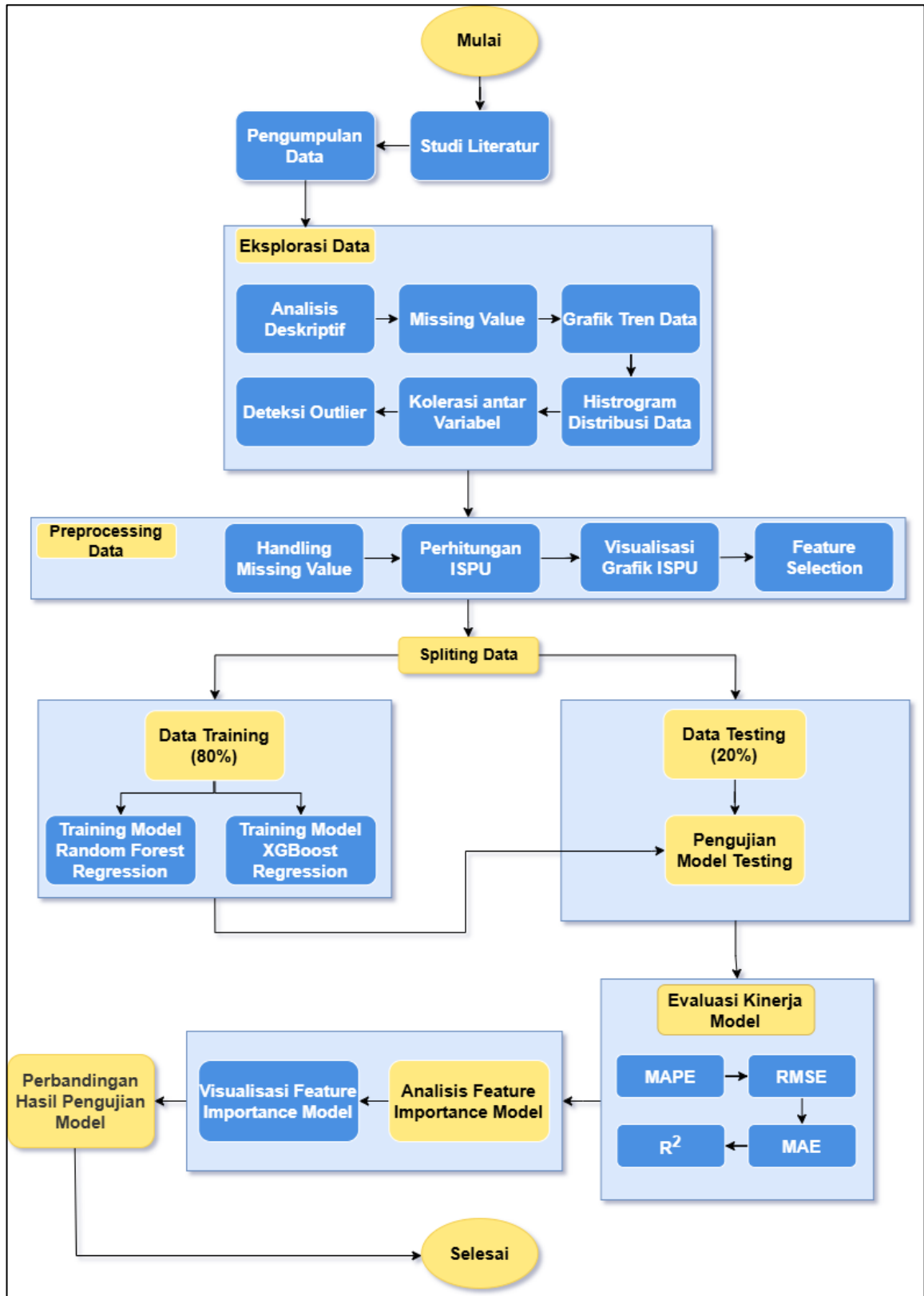
3.2.2 Perangkat Lunak

- A. Sistem Operasi: *Windows 11 Home Single Language*
- B. *Google Colab*
- C. *Microsoft Word 365*
- D. *Microsoft Excel 365*
- E. *Draw.io*

3.3 Tahapan Penelitian

Dalam penelitian ini terdapat beberapa tahapan yang dilakukan secara sistematis untuk mencapai tujuan penelitian. Tahapan tersebut meliputi proses pengumpulan data, eksplorasi data, *preprocessing* data, pembagian data (*splitting data*), pelatihan model (*training model*), pengujian model (*testing model*), evaluasi kinerja model, serta analisis dan perbandingan hasil model. Untuk memberikan gambaran yang lebih jelas mengenai alur penelitian yang dilakukan,

Tahapan-tahapan tersebut disajikan dalam bentuk diagram alur pada Gambar 8.



Gambar 8. Alur Tahapan Penelitian

Berdasarkan Gambar 8, tahapan penelitian diawali dengan studi literatur dan pengumpulan data. Studi literatur dilakukan untuk memperoleh dasar teori yang relevan, sedangkan pengumpulan data bertujuan untuk mendapatkan data yang akan digunakan dalam penelitian.

Tahap berikutnya adalah eksplorasi data yang bertujuan untuk memahami karakteristik data. Pada tahap ini dilakukan analisis deskriptif untuk melihat gambaran umum data, identifikasi *missing value*, visualisasi grafik tren data, deteksi *outlier*, analisis korelasi antar variabel, serta distribusi data menggunakan histogram.

Tahap selanjutnya adalah *preprocessing* data yang meliputi penanganan *missing value*, perhitungan nilai ISPU, visualisasi data, serta pemilihan fitur (*feature selection*). Tahap ini bertujuan untuk mempersiapkan data agar siap digunakan dalam proses pemodelan.

Data kemudian dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih model menggunakan algoritma *Random Forest Regression* dan *XGBoost Regression*, sedangkan data *testing* digunakan untuk menguji kinerja model yang telah dibangun.

Proses berikutnya adalah evaluasi kinerja model menggunakan beberapa metrik evaluasi. Selain itu, dilakukan analisis *feature importance* untuk mengetahui pengaruh masing-masing variabel terhadap model, yang kemudian divisualisasikan untuk mempermudah interpretasi.

Tahap akhir dalam penelitian ini adalah perbandingan hasil pengujian model untuk menentukan model terbaik.

3.3.1 Studi Literatur

Tahap awal dalam penelitian ini adalah melakukan studi literatur dengan mencari dan mempelajari berbagai referensi yang relevan dengan topik penelitian. Referensi yang digunakan berupa jurnal, buku, dan sumber ilmiah lainnya yang berkaitan dengan kualitas udara, metode *machine learning*, serta algoritma *Random Forest* dan *XGBoost*.

Melalui studi literatur ini, peneliti memahami konsep dasar yang digunakan dalam penelitian, seperti karakteristik data kualitas udara, perhitungan Indeks Standar Pencemar Udara (ISPU), serta metode evaluasi model seperti RMSE, MAE, MAPE, dan R^2 . Hasil dari studi literatur ini kemudian digunakan sebagai dasar dalam menentukan metode dan tahapan penelitian yang akan dilakukan.

3.3.2 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan memperoleh data kualitas udara dari stasiun pemantauan *Automatic Air Quality Monitoring System* (AQMS) yang dikelola oleh Dinas Lingkungan Hidup Kota Bandar Lampung. Stasiun pemantauan tersebut berlokasi di wilayah Sukarame dan menjadi acuan utama dalam penelitian ini.

Data yang digunakan mencakup periode 2023 hingga 2025 dengan interval waktu harian. Parameter yang digunakan meliputi beberapa jenis polutan udara, yaitu PM_{10} , $PM_{2,5}$, SO_2 , CO, O_3 , dan NO_2 , serta parameter meteorologi berupa kelembaban, kecepatan angin, dan suhu.

Data yang diperoleh dari stasiun AQMS merepresentasikan kondisi kualitas udara di wilayah sekitar lokasi pengamatan dengan jangkauan pemantauan kurang lebih ± 5 km yang dipengaruhi oleh kondisi meteorologi. Oleh karena itu, data yang digunakan dalam penelitian ini dapat menggambarkan kondisi kualitas udara di wilayah Sukarame dan sekitarnya.

Seluruh data yang telah dikumpulkan kemudian disusun dalam bentuk dataset dan disimpan dalam format Microsoft Excel untuk memudahkan proses pengolahan data. Dataset ini bersifat *time series* karena tersusun berdasarkan urutan waktu pengamatan secara harian.

Data yang telah dikumpulkan selanjutnya digunakan sebagai dasar dalam tahap eksplorasi data untuk memahami karakteristik data, kemudian dilanjutkan dengan proses *preprocessing* sebelum dilakukan pemodelan menggunakan algoritma *Random Forest* dan *XGBoost*.

3.3.3 Eksplorasi Data

Eksplorasi data merupakan tahap awal dalam penelitian yang dilakukan untuk memahami karakteristik dataset sebelum memasuki proses *preprocessing* dan pemodelan. Tahap ini penting dilakukan karena data yang digunakan memiliki berbagai variabel dengan karakteristik yang berbeda, sehingga perlu dianalisis terlebih dahulu untuk mengetahui kondisi awal data.

Dataset yang digunakan berupa data deret waktu (*time series*) dengan interval harian yang memungkinkan adanya variasi nilai, ketidaklengkapan data, maupun nilai ekstrem. Oleh karena itu, diperlukan proses eksplorasi data untuk mengidentifikasi potensi permasalahan yang dapat memengaruhi proses analisis dan pemodelan.

Pada penelitian ini, eksplorasi data dilakukan terhadap dataset kualitas udara yang terdiri dari variabel polutan, yaitu PM_{10} , $PM_{2,5}$, SO_2 , CO , O_3 , dan NO_2 , serta variabel meteorologi berupa kelembaban, kecepatan angin, dan suhu.

Tahapan eksplorasi data dilakukan secara sistematis sesuai dengan alur penelitian yang disajikan pada Gambar 6, yang meliputi analisis deskriptif, *missing value*, grafik tren data, distribusi data menggunakan histogram, analisis korelasi antar variabel, serta deteksi *outlier*. Melalui tahapan ini, diharapkan diperoleh pemahaman yang komprehensif mengenai kondisi data sehingga dapat menentukan langkah pengolahan data yang tepat pada tahap selanjutnya.

A. Analisis Deskriptif

Pada tahap awal, data yang telah dikumpulkan perlu dipahami karakteristik dasarnya untuk mengetahui kondisi awal dataset. Hal ini penting karena setiap variabel memiliki rentang nilai dan tingkat variasi yang berbeda.

Oleh karena itu, dilakukan analisis deskriptif untuk memberikan gambaran umum mengenai karakteristik data. Analisis ini mencakup

perhitungan statistik dasar seperti jumlah data, nilai minimum, maksimum, rata-rata (mean), median, serta standar deviasi untuk setiap variabel.

Analisis dilakukan terhadap seluruh variabel, baik variabel polutan maupun variabel meteorologi. Melalui tahap ini, peneliti dapat memahami karakteristik awal data serta mengidentifikasi kondisi umum dari masing-masing variabel sebelum dilakukan proses pengolahan lebih lanjut.

B. *Missing Value*

Setelah memahami karakteristik dasar data, langkah selanjutnya adalah memeriksa kelengkapan data. Hal ini penting karena data yang tidak lengkap dapat memengaruhi hasil analisis dan pemodelan.

Pada tahap ini dilakukan *missing value* untuk mengetahui apakah terdapat data yang tidak terisi pada dataset. *Missing value* dapat terjadi akibat gangguan alat pengukuran, keterbatasan proses pencatatan data, atau ketidaksesuaian waktu pengambilan data.

Pemeriksaan dilakukan terhadap seluruh variabel dalam dataset sehingga dapat diketahui jumlah dan persebaran data yang hilang pada masing-masing variabel. Melalui tahap ini, peneliti dapat mengetahui kondisi kelengkapan data serta menentukan langkah penanganan yang tepat pada tahap *preprocessing*.

C. Grafik Tren Data

Data kualitas udara dan meteorologi yang digunakan dalam penelitian ini bersifat deret waktu, sehingga penting untuk memahami pola perubahan data terhadap waktu.

Oleh karena itu, dilakukan visualisasi grafik tren data untuk melihat pola perubahan nilai setiap variabel berdasarkan waktu. Grafik disajikan dalam bentuk deret waktu (*time series*) menggunakan grafik garis, di mana setiap variabel ditampilkan dalam subplot yang berbeda.

Melalui visualisasi ini, seluruh variabel seperti PM₁₀, PM_{2,5}, SO₂, CO, O₃, NO₂, kelembaban, kecepatan angin dan suhu dapat diamati secara bersamaan. Tahap ini dilakukan untuk memahami pola perubahan data dari waktu ke waktu serta memberikan gambaran awal mengenai dinamika data sebelum dilakukan proses pengolahan lebih lanjut.

D. Histogram Distribusi Data

Untuk melihat pola perubahan terhadap waktu, penting juga memahami distribusi nilai dari setiap variabel. Oleh karena itu, dilakukan analisis distribusi data menggunakan histogram untuk melihat penyebaran nilai pada masing-masing variabel dalam dataset. Histogram disajikan untuk seluruh variabel, baik variabel polutan maupun variabel meteorologi.

Melalui histogram, dapat diketahui bagaimana distribusi nilai pada setiap variabel, apakah memiliki kecenderungan tertentu atau penyebaran yang merata. Tahap ini membantu peneliti dalam memahami karakteristik distribusi data sebagai dasar dalam menentukan langkah pengolahan data pada tahap *preprocessing*.

E. Korelasi Antar Variabel

Variabel dalam dataset tidak berdiri sendiri, melainkan dapat saling berhubungan satu sama lain. Oleh karena itu, penting untuk mengetahui hubungan antar variabel dalam dataset. Pada tahap ini dilakukan analisis korelasi untuk mengetahui hubungan antar variabel, baik antar variabel polutan maupun antara variabel polutan dan variabel meteorologi.

Hasil analisis korelasi divisualisasikan dalam bentuk heatmap sehingga memudahkan dalam melihat tingkat keterkaitan antar variabel. Melalui tahap ini, peneliti dapat memahami hubungan antar variabel yang nantinya digunakan sebagai dasar dalam proses pemilihan fitur (*feature selection*) pada tahap pemodelan.

F. Deteksi *Outlier*

Selain kelengkapan dan distribusi data, keberadaan nilai ekstrem juga perlu diperhatikan karena dapat memengaruhi hasil analisis. Oleh karena

itu, dilakukan deteksi outlier untuk mengidentifikasi nilai yang menyimpang dari pola umum dalam dataset. Pemeriksaan dilakukan terhadap seluruh variabel menggunakan visualisasi boxplot untuk melihat sebaran data.

Melalui tahap ini, peneliti dapat mengetahui keberadaan nilai ekstrem pada setiap variabel yang berpotensi memengaruhi proses analisis dan pemodelan. Hasil dari tahap ini digunakan sebagai pertimbangan dalam proses *preprocessing* agar data yang digunakan lebih akurat dan representatif.

3.3.4 *Preprocessing*

Preprocessing data merupakan tahap yang dilakukan untuk menyiapkan dataset agar dapat digunakan dalam proses pemodelan. Tahap ini memiliki peran penting karena kualitas data yang digunakan akan sangat memengaruhi hasil analisis dan performa model yang dibangun.

Berdasarkan hasil eksplorasi data pada tahap sebelumnya, data yang digunakan masih memiliki beberapa permasalahan, seperti adanya nilai yang belum lengkap (*missing value*) serta perlunya transformasi data dari nilai konsentrasi polutan menjadi bentuk indeks kualitas udara. Permasalahan tersebut perlu ditangani agar data yang digunakan menjadi lebih representatif.

Dataset yang digunakan merupakan data deret waktu (*time series*) dengan interval harian. Karakteristik ini menunjukkan bahwa setiap data memiliki keterkaitan terhadap waktu, sehingga proses pengolahan perlu dilakukan dengan pendekatan yang sesuai agar pola temporal dalam data tetap terjaga.

Oleh karena itu, pada tahap *preprocessing* dilakukan beberapa proses yang meliputi penanganan *missing value*, perhitungan Indeks Standar Pencemar Udara (ISPU), visualisasi data, serta pemilihan fitur (*feature selection*). Tahapan ini bertujuan untuk menghasilkan data yang lebih siap digunakan dalam proses pemodelan.

A. *Handling Missing Value*

Pada tahap ini dilakukan penanganan terhadap data yang belum lengkap (*missing value*) agar seluruh data dapat digunakan dalam proses analisis dan pemodelan.

Keberadaan *missing value* pada dataset dapat menyebabkan informasi yang diperoleh menjadi tidak lengkap serta berpotensi memengaruhi kinerja model. Oleh karena itu, diperlukan metode yang mampu mengisi nilai yang hilang tanpa mengubah karakteristik data secara signifikan.

Karena dataset yang digunakan merupakan data deret waktu (*time series*) dengan interval harian, maka digunakan metode interpolasi linear berbasis waktu. Metode ini digunakan untuk memperkirakan nilai yang hilang dengan mempertimbangkan pola perubahan data pada waktu sebelumnya dan sesudahnya.

Metode *interpolasi linear* yang telah dijelaskan pada BAB 2 digunakan dalam penelitian ini dengan persamaan sebagai berikut:

$$f_1 = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0) \quad (12)$$

Sumber : (Kusumawardana, 2021)

Keterangan:

- x = titik atau variabel yang akan dicari nilainya
- x_0 = titik pertama yang diketahui
- x_1 = titik kedua yang diketahui
- $f(x_0)$ = nilai fungsi pada titik x_0
- $f(x_1)$ = nilai fungsi pada titik x_1
- $f(x)$ = nilai hasil *interpolasi*

Proses *interpolasi* dilakukan dengan memanfaatkan informasi waktu sehingga nilai yang hilang dapat diestimasi secara berurutan. Proses ini dilakukan sebelum tahap pembagian data (*data splitting*) sehingga hubungan temporal antar data tetap terjaga. Dengan demikian, data menjadi lebih lengkap dan tetap mempertahankan pola temporal yang terdapat pada dataset. Tahap ini bertujuan untuk memastikan bahwa

dataset yang digunakan tidak memiliki nilai kosong sehingga dapat digunakan secara optimal pada tahap selanjutnya.

B. Perhitungan Indeks Standar Pencemar Udara (ISPU)

Setelah data dinyatakan lengkap, dilakukan proses transformasi dari nilai konsentrasi polutan menjadi Indeks Standar Pencemar Udara (ISPU). Transformasi ini diperlukan karena setiap parameter polutan memiliki satuan yang berbeda, sehingga perlu dikonversi ke dalam bentuk indeks yang seragam agar lebih mudah dianalisis.

Perhitungan ISPU pada penelitian ini mengacu pada Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor 14 Tahun 2020. Parameter yang digunakan dalam penelitian ini meliputi PM₁₀, PM_{2,5}, SO₂, CO, O₃, dan NO₂.

Perhitungan nilai ISPU dilakukan dengan menggunakan persamaan sebagai berikut:

$$I = \frac{(I_a - I_b)}{(\chi_a - \chi_b)} \times (\chi_x - \chi_b) + I_b \quad (13)$$

Sumber : (Permen LHK, 2020)

Keterangan:

- I = ISPU terhitung
- I_a = ISPU batas atas
- I_b = ISPU batas bawah
- χ_a = konsentrasi ambien batas atas ($\mu\text{g}/\text{m}^3$)
- χ_b = konsentrasi ambien batas bawah ($\mu\text{g}/\text{m}^3$)
- χ_x = konsentrasi ambien nyata hasil pengukuran ($\mu\text{g}/\text{m}^3$)

Persamaan tersebut digunakan untuk mengonversi nilai konsentrasi masing-masing parameter menjadi nilai ISPU berdasarkan rentang yang telah ditentukan. Selanjutnya, nilai ISPU dihitung untuk setiap parameter polutan. Nilai ISPU akhir ditentukan dengan mengambil nilai tertinggi dari seluruh parameter pada waktu tertentu, sehingga dapat merepresentasikan kondisi kualitas udara secara keseluruhan.

Nilai ISPU yang dihasilkan kemudian digunakan sebagai variabel target dalam proses pemodelan, sedangkan parameter pencemar udara digunakan sebagai variabel input (fitur) dalam model regresi. Tahap ini bertujuan untuk menghasilkan variabel target yang akan digunakan dalam proses pemodelan.

C. Visualisasi Grafik ISPU

Setelah nilai ISPU diperoleh, dilakukan visualisasi data untuk melihat pola perubahan kualitas udara terhadap waktu. Visualisasi disajikan dalam bentuk grafik garis berbasis deret waktu, dengan waktu sebagai sumbu horizontal dan nilai ISPU sebagai sumbu vertikal. Melalui visualisasi ini, pola perubahan kualitas udara dapat diamati secara umum.

Tahap ini bertujuan untuk memberikan gambaran awal mengenai dinamika kualitas udara sebelum digunakan dalam proses pemodelan.

D. Feature Selection

Tahap selanjutnya adalah pemilihan fitur (*feature selection*), yaitu menentukan variabel yang akan digunakan sebagai input dalam model. Pemilihan fitur dilakukan karena tidak semua variabel dalam dataset memiliki kontribusi yang sama terhadap variabel target. Oleh karena itu, diperlukan pemilihan variabel yang relevan agar model dapat bekerja dengan lebih efektif.

Variabel yang digunakan sebagai fitur meliputi parameter polutan, yaitu PM₁₀, PM_{2,5}, SO₂, CO, O₃, dan NO₂, serta parameter meteorologi seperti kelembaban, kecepatan angin dan suhu. Sementara itu, nilai ISPU digunakan sebagai variabel target.

Tahap ini bertujuan untuk memastikan bahwa model menggunakan variabel yang sesuai sehingga dapat meningkatkan kinerja model dalam melakukan analisis.

3.3.5 *Splitting Data*

Tahap *splitting data* merupakan proses pembagian dataset menjadi dua bagian, yaitu *data training* dan *data testing*, yang digunakan dalam proses analisis menggunakan metode *machine learning*. Dataset yang digunakan dalam penelitian ini merupakan data kualitas udara Kota Bandar Lampung periode 2023–2025 yang terdiri dari parameter polutan dan parameter meteorologi.

Pembagian data dilakukan dengan tujuan agar model dapat mempelajari pola hubungan antar variabel dari sebagian data, kemudian kinerjanya dapat dievaluasi menggunakan data lain yang tidak digunakan pada saat pelatihan model. Dengan demikian, hasil evaluasi yang diperoleh dapat menggambarkan kemampuan model dalam merepresentasikan pola data secara umum.

Pada penelitian ini, pembagian data dilakukan menggunakan metode *train-test split* dengan perbandingan 80% data *training* dan 20% data *testing*. Selain itu, data dibagi secara acak (*shuffle=True*) untuk memastikan distribusi data pada masing-masing bagian menjadi lebih merata.

Penggunaan pembagian data secara acak dilakukan karena penelitian ini tidak berfokus pada peramalan berbasis waktu (*time series forecasting*), melainkan pada analisis hubungan antar variabel dalam periode data yang sama (2023–2025). Oleh karena itu, pengacakan data membantu model dalam mempelajari variasi data secara lebih menyeluruh.

A. *Data Training*

Data *training* merupakan bagian dari dataset yang digunakan untuk melatih model dalam mengenali pola hubungan antara variabel input dan variabel target. Variabel input dalam penelitian ini terdiri dari parameter polutan udara (PM₁₀, PM_{2,5}, SO₂, CO, O₃, dan NO₂) serta parameter

meteorologi (kelembaban, kecepatan angin, dan suhu), sedangkan variabel target yang digunakan adalah nilai ISPU.

Sebanyak 80% dari total data digunakan sebagai data *training*. Pada tahap ini dilakukan proses pelatihan model menggunakan dua algoritma, yaitu *Random Forest Regression* dan *XGBoost Regression*. Kedua model tersebut digunakan untuk mempelajari hubungan antara parameter pencemar udara dan kondisi meteorologi terhadap nilai ISPU. Proses pelatihan ini bertujuan untuk membentuk model yang mampu merepresentasikan pola data yang terdapat dalam dataset.

B. Data Testing

Data *testing* merupakan bagian dari dataset yang digunakan untuk mengevaluasi kinerja model setelah proses pelatihan selesai. Data ini tidak dilibatkan dalam proses pelatihan sehingga dapat digunakan untuk menilai kemampuan model dalam mengenali pola pada data yang belum dipelajari sebelumnya.

Sebanyak 20% dari total data digunakan sebagai data *testing*. Pada tahap ini, model *Random Forest* dan *XGBoost* yang telah dilatih diterapkan pada data *testing*. Hasil yang diperoleh kemudian dievaluasi menggunakan metrik MAE, RMSE, MAPE, dan R^2 untuk menilai performa masing-masing model dalam menganalisis kualitas udara.

Berdasarkan hasil evaluasi tersebut, dapat diketahui tingkat akurasi masing-masing model serta dilakukan perbandingan untuk menentukan model dengan performa terbaik.

3.3.6 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk menilai kemampuan model *Random Forest* dan *XGBoost* dalam menganalisis kualitas udara berdasarkan data yang digunakan. Pada tahap ini, model terlebih dahulu dilatih menggunakan data *training*, kemudian dilakukan pengujian menggunakan data *testing*. Hasil keluaran model selanjutnya

dibandingkan dengan data aktual untuk mengetahui tingkat kesesuaian antara keduanya.

Pada tahap ini, selanjutnya dilakukan perhitungan menggunakan beberapa metrik evaluasi, yaitu MAPE digunakan untuk mengukur persentase kesalahan, RMSE digunakan untuk melihat besarnya kesalahan secara lebih sensitif, MAE digunakan untuk mengetahui rata-rata kesalahan, dan R^2 digunakan untuk melihat kemampuan model dalam menggambarkan pola data.

Selain itu, dilakukan juga perbandingan hasil pada *data training* dan *data testing* untuk melihat konsistensi kinerja model. Jika perbedaan nilainya terlalu besar, maka model dapat dikatakan kurang mampu bekerja dengan baik pada data yang digunakan. Hasil evaluasi ini digunakan untuk mengetahui performa masing-masing model dalam menganalisis kualitas udara.

3.3.7 Analisis *Feature Importance*

Analisis *feature importance* dilakukan untuk mengetahui seberapa besar pengaruh masing-masing variabel terhadap model dalam menganalisis kualitas udara. Variabel yang digunakan meliputi parameter polutan udara dan parameter meteorologi. Pada tahap ini, model *Random Forest* dan *XGBoost* digunakan untuk menghitung tingkat kepentingan setiap variabel berdasarkan kontribusinya dalam proses pembentukan model. Nilai *feature importance* yang diperoleh menunjukkan seberapa besar peran setiap variabel dalam membantu model memahami pola data.

Nilai tersebut kemudian diurutkan dari yang terbesar hingga yang terkecil dan divisualisasikan dalam bentuk grafik untuk memudahkan interpretasi. Variabel dengan nilai *feature importance* yang lebih tinggi menunjukkan pengaruh yang lebih besar dibandingkan variabel lainnya. Melalui analisis ini, dapat diketahui variabel yang paling dominan dalam analisis kualitas udara di Kota Bandar Lampung.

3.3.8 Perbandingan Hasil Pengujian Model

Perbandingan hasil pengujian model dilakukan untuk menentukan model yang memiliki performa terbaik dalam menganalisis kualitas udara. Proses ini dilakukan dengan membandingkan hasil evaluasi dari model *Random Forest* dan *XGBoost* berdasarkan metrik MAPE, RMSE, MAE, dan R^2 .

Model dengan nilai MAPE, RMSE, dan MAE yang lebih kecil serta nilai R^2 yang lebih mendekati 1 menunjukkan bahwa model tersebut lebih baik dalam menggambarkan kondisi data kualitas udara.

Hasil pada data training dan data testing juga dibandingkan untuk melihat kestabilan model. Model yang memiliki perbedaan nilai yang tidak terlalu besar menunjukkan bahwa model mampu bekerja dengan baik pada data yang digunakan.

Berdasarkan hasil perbandingan tersebut, dapat ditentukan model yang paling sesuai digunakan dalam analisis kualitas udara di Kota Bandar Lampung.

V. SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai evaluasi kinerja model *Random Forest* dan *XGBoost* dalam analisis kualitas udara di Kota Bandar Lampung, diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil evaluasi kinerja algoritma, model *Random Forest* dan *XGBoost* menunjukkan performa yang sangat baik dalam analisis kualitas udara di Kota Bandar Lampung. Hasil pengujian menunjukkan bahwa model *Random Forest* memperoleh nilai koefisien determinasi (R^2) sebesar 0,99, sedangkan model *XGBoost* memperoleh nilai R^2 sebesar 0,98. Nilai tersebut menunjukkan bahwa kedua model mampu merepresentasikan variasi data kualitas udara dengan tingkat akurasi yang sangat tinggi. Berdasarkan data kualitas udara yang dikumpulkan dari Kecamatan Sukarame, Kota Bandar Lampung, dengan radius pengamatan sekitar ± 5 km dari lokasi alat, algoritma *Random Forest* menunjukkan performa yang sedikit lebih baik dibandingkan *XGBoost*. Hasil ini menunjukkan bahwa *Random Forest* lebih efektif dalam menganalisis kualitas udara.
2. Berdasarkan hasil perbandingan kinerja model menggunakan metrik evaluasi MAPE, MAE, RMSE, dan R^2 , model *Random Forest* menunjukkan performa yang lebih baik dibandingkan *XGBoost*. Hal ini ditunjukkan oleh nilai MAPE sebesar 1,178%, RMSE sebesar 1,994, dan MAE sebesar 0,743 yang lebih rendah dibandingkan dengan *XGBoost* yang memiliki nilai MAPE sebesar 1,570%, RMSE sebesar 2,074 dan MAE sebesar 0,969. Selain itu, nilai R^2 pada *Random Forest* juga sedikit lebih tinggi dibandingkan *XGBoost*. Perbedaan nilai evaluasi antara kedua

model relatif kecil, sehingga menunjukkan bahwa kedua model memiliki kinerja yang hampir setara dalam menganalisis kualitas udara. Namun demikian, *Random Forest* tetap memberikan performa yang lebih baik secara keseluruhan.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, maka saran yang dapat diberikan adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk mengembangkan variabel yang digunakan dengan menambahkan faktor lain yang belum digunakan dalam penelitian ini, seperti curah hujan atau variabel lingkungan lainnya, sehingga analisis kualitas udara dapat dilakukan secara lebih komprehensif.
2. Selain itu, penelitian selanjutnya dapat mengembangkan metode yang digunakan, seperti penerapan model *deep learning*, optimasi *hyperparameter* secara lebih mendalam, serta penerapan *Genetic Algorithm* untuk mengoptimalkan metode *boosting*, sehingga kinerja model dalam analisis kualitas udara dapat ditingkatkan, terutama dalam menangani pola data yang lebih kompleks.

DAFTAR PUSTAKA

- Akbar, M. R., Akbar, M. I., Darajatun, A., Hs, J., & Waluyo, R. (2022). Analisis Regulasi Uji Emisi Gas Buang Kendaraan Berdasarkan Pengaruhnya Terhadap Indeks Kualitas Udara di DKI Jakarta Menggunakan Metode Korelasi Pearson dan Regresi Linear. *Jurnal Teknologi Lingkungan*, 15(1), 137–146. <https://doi.org/10.36456/jstat.vol15.no1.a5216>
- Ali, M. C., Ebrahim, E. E. M., & Abonazel, M. R. (2025). Air quality forecasting using a modified statistical approach: Combining statistical and machine learning methods. *International Journal of Innovative Research and Scientific Studies*, 8(4), 1321–1335. <https://doi.org/10.53894/ijriss.v8i4.8061>
- Aminudin, C., Fadly, R., & Hermawan. (2021). Interaksi Hukum dan Sains dalam Siklus Kebijakan Baku Mutu Udara Ambien. *Jurnal Ekologi, Masyarakat & Sains*, 2(2), 54–59. <https://doi.org/10.55448/ems.v2i2.47>
- Arifianti, F. P., & Salam, A. (2024). *XGBoost and Random Forest Optimization using SMOTE to Classify Air Quality*. 6(1), 1–8.
- Ayus, I., Natarajan, N., & Gupta, D. (2023). Comparison of machine learning and deep learning techniques for the prediction of air pollution: a case study from China. *Asian Journal of Atmospheric Environment*, 17(4). <https://doi.org/10.1007/s44273-023-00005-w>
- Azizah, D. N., Heranurweni, S., Ode, L., & Idris, M. (2025). Internet of Things Based Air Quality Monitoring System with Automatic Notification. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 776–787. <https://doi.org/10.57152/malcom.v5i3.1945>
- Bai, X., Zhang, N., Cao, X., & Chen, W. (2024). Prediction of PM2.5 concentration based on a CNN-LSTM neural network algorithm. *PeerJ*, e17811. <https://doi.org/10.7717/peerj.17811>
- Bhathal, B. S., & Gupta, G. (2021). Air quality prediction using a machine learning hybrid model in Punjab. *Journal of Electrical Systems*, 1, 89–106. <https://journal.esrgroups.org/jes/article/view/7362>
- Binbusayyis, A., Attique, M., Mohamed, K., Ahmed, M., & Emmanuel, A. W. R. S. (2024). A deep learning approach for prediction of air quality index in smart city. *Discover Sustainability*, 5, 89. <https://doi.org/10.1007/s43621-024-00272-9>
- Cahyani, N. D. (2024). *Klasifikasi stunting di Kabupaten Lampung Barat berdasarkan data antropometri menggunakan analisis algoritma support vector machine dan extreme gradient boosting*. Universitas Lampung. <https://digilib.unila.ac.id/85670/>

- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 1–24. <https://doi.org/10.7717/peerj-cs.623>
- Christian, M. M., & Choi, H. (2024). Air quality forecasting using machine learning: A global perspective with relevance to low-resource settings. *SIBR 2024 Conference on Interdisciplinary Business and Economics Research*, 1–16. https://buscompress.com/uploads/3/4/9/8/34980536/riber_13-4_40_k24-155_605-619.pdf
- Endra, R. Y., & Saputra, G. A. (2022). Implementasi fuzzy inference system (FIS) metode Tsukamoto untuk monitoring kualitas udara. *Jurnal Komputasi*, 10(1).
- Frommeyer, T. C., Gilbert, M. M., Fursmidt, R. M., Park, Y., Khouzam, J. P., Brittain, G. V., Frommeyer, D. P., Bett, E. S., & Bihl, T. J. (2025). Reinforcement learning and its clinical applications within healthcare: A systematic review of precision medicine and dynamic treatment regimes. *Healthcare*, 13, 1752. <https://doi.org/10.3390/healthcare13141752>
- Golec, M., Sahin, E., & Singh, S. (2025). Artificial Intelligence (AI): Foundations, trends and future directions. *Telematics and Informatics Reports*, 20, 100265. <https://doi.org/10.1016/j.teler.2025.100265>
- Hasnataeni, Y., Saefuddin, A., & Soleh, A. M. (2025). Comparison of ensemble forest-based methods performance for imbalanced data classification. *Scientific Journal of Informatics*, 12(2), 183–198.
- Iqbal, M., Hermanto, B., Febriansyah, F. E., & Ridho, M. (2019). Sistem Pendeteksi Polusi Udara di Kota Bandar Lampung Menggunakan Teknologi Internet of Things (IoT). *Jurnal Komputasi*, 7(2).
- Kiswandono, A. A., Aprilia, K. C., & Rahmawati, A. (2025). Analysis of Nitrogen Dioxide as an Air Pollutant in Office , Industrial , Residential , and Transportation Areas in Lampung Province. *Jurnal Kimia Dan Pendidikan Kimia*, 10(1), 1–17. <https://doi.org/10.20961/jkpk.v10i1.92581>
- Komarudin, J. H., Apriyana, M., Ergantara, R. I., & Nasoetion, P. (2023). Analisis emisi karbon monoksida akibat kemacetan kendaraan di Kota Bandar Lampung (studi kasus: palang pintu perlintasan kereta api). *Jurnal Serambi Engineering*, 8(3), 6573–6581. <https://doi.org/10.32672/jse.v8i3.5512>
- Kurniasari, D., Rafdiono, M. N. A., & Warsono. (2024). Analisis Struktur Terbaik Neural Network dengan Algoritma Backpropagation dalam Memprediksi Indeks Kandungan Sulfida (SO₂) di Ibu Kota Jakarta. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 12(2), 321–329.
- Lei, T. M. T., Ng, S. C. W., & Siu, S. W. I. (2023). Application of ANN , XGBoost , and Other ML Methods to Forecast Air Quality in Macau. *Sustainability*, 15. <https://doi.org/10.3390/su15065341>

- Mahbodi, M., & Karasfi, B. (2025). Forecasting the Air Quality Index Using Machine Learning Models, Bayesian Optimization, and the Development of the S-GBR Model Incorporating Seasonal Variables. *Journal of Resource Management and Decision Engineering*, 4(4), 1–16.
- Mampitiya, L., Rathnayake, N., Leon, L. P., Mandala, V., Azamathulla, H., Shelton, S., Hoshino, Y., & Rathnayake, U. (2023). Machine Learning Techniques to Predict the Air Quality Using Meteorological Data in Two Urban Areas in Sri Lanka. *Environments*, 10, 1–18.
- Manik, A. R., Syahri, A., & Perdana, A. (2026). Optimasi Random Forest Menggunakan Genetic Algorithm untuk Klasifikasi Kualitas Udara Berdasarkan Data ISPU (Indeks Standar Pencemaran Udara) Random Forest Optimization Using Genetic Algorithm for Air Quality Classification Based on ISPU (Air Pollution. *Jurnal Intelek Dan Cendekiawan Nusantara*, 2(6), 11987–12000.
- Manisalidis, I., Stavropoulou, E., & Stavropoulos, A. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, 8, 1–13. <https://doi.org/10.3389/fpubh.2020.00014>
- Muhammad, R., Latif, A., Iqbal, T., Qader, I. A., Ikram, A., Alsolai, H., Alabdullah, B., Alhayan, F., & Ghazal, T. M. (2025). Interpretable machine learning framework for predicting Urban air quality. *PLOS ONE*, 20, 1–21. <https://doi.org/10.1371/journal.pone.0336241>
- Naizabayeva, L., Sembina, G., Aliman, A., & Satymbekov, M. (2025). Air Pollution Forecasting in Almaty using Ensemble Machine Learning Models. *Journal of Applied Data Sciences*, 6(4), 2461–2476.
- Natarajan, S. K., Shanmurthy, P., Arockiam, D., Balusamy, B., & Selvarajan, S. (2024). Optimized machine learning model for air quality index prediction in major cities in India. *Scientific Reports*, 14, 1–18. <https://doi.org/10.1038/s41598-024-54807-1>
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72. <https://doi.org/10.54082/jiki.168>
- Özüpak, Y., Alpsalaz, F., & Aslan, E. (2025). Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction. *Water, Air, & Soil Pollution*, 236(7), 1–17. <https://doi.org/10.1007/s11270-025-08122-8>
- Patel, P., Patel, S., Shah, K., Desai, K., Patel, S., & Shah, M. (2025). A systematic study on PM2.5 and PM10 concentration prediction in air pollution using machine learning and deep learning models. *Environmental Chemistry and Ecotoxicology*, 7, 1401–1415. <https://doi.org/10.1016/j.eneco.2025.07.001>
- Permen LHK. (2020). *Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020 tentang Indeks Standar Pencemar Udara*.

- Rahmawati, V., Hayat, A. L., & Salam, A. (2024). Analisis dampak pencemaran udara terhadap kesehatan masyarakat di perkotaan. *Jurnal Sosial Dan Pengabdian Masyarakat*, 2(3), 145.
- Rajagukguk, I. M., Rudy, H., Julian, & Halim, R. (2025). Comparative Analysis of XGBoost, Random Forest, and Logistic Regression for Classifying Jakarta's Air Pollution Index (ISPU). *Procedia Computer Science*, 269, 108–120. <https://doi.org/10.1016/j.procs.2025.08.264>
- Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the Indian coastal city of Visakhapatnam. *Chemosphere*, 338, 139518. <https://doi.org/10.1016/j.chemosphere.2023.139518>
- Razzaq, K., & Shah, M. (2025). Machine Learning and Deep Learning Paradigms : From Techniques to Practical Applications and Research Frontiers. *Computers*, 14, 145. <https://doi.org/10.3390/computers14030093>
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python. *Bina Insani ICT Journal*, 7(2), 156–165. <https://doi.org/10.51211/biict.v7i2.1422>
- Rivaldi, A., Ramadhani, A., Ramadhan, I., & Nuryamin, Y. (2025). Prediksi dan Pemodelan Kualitas Udara Menggunakan Random Forest dan Gradient Boosting Jakarta dan Tangerang. *Jurnal Nasional Teknologi Komputer*, 5(4), 1149–1157.
- Saleh, K., Hameedy, M. A. El, Metwally, A., & Mabrouk, W. M. (2025). XGBoost regression for robust acoustic impedance prediction in the absence of density and sonic logs. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-025-24727-9>
- Sari, E. M., Sabila, C., Adam, R. F., & Kurniawan, R. (2025). Analisis dan Prediksi Indeks Kualitas Udara Jakarta: Penerapan Algoritma XGBoost. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(2), 161–169. <https://doi.org/10.25077/TEKNOSI.v11i2.2025.161-169>
- Setiawan, A., Wibowo, U. L., Mubarak, A., Larasati, K., & Hammad, J. A. H. (2024). Random Forest Algorithm to Measure the Air Pollution Standard Index. *Knowledge Engineering and Data Science*, 7(1), 86–100. <https://doi.org/10.17977/um018v7i12024p86-100>
- Shakya, K. S., Alavi, A., Porteous, J., Priti, P., Laddi, A., & Jaiswal, M. (2024). A Critical Analysis of Deep Semi-Supervised Learning Approaches for Enhanced Medical Image Classification. *Information*, 15(2), 89–110. <https://doi.org/10.3390/info15050246>
- Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning : A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(07), 3898–3904.
- Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence , machine learning and deep learning in advanced robotics , a review. *Cognitive Robotics*, 3, 54–70. <https://doi.org/10.1016/j.cogr.2023.04.001>

- Tırnk, S. (2025). Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM, and SVM. *PLOS ONE*, 20(4), 1–21.
- Yenkikar, A., Mishra, V. P., Bali, M., & Ara, T. (2025). Explainable forecasting of air quality index using a hybrid random forest and ARIMA model. *MethodsX*, 15(July), 103517. <https://doi.org/10.1016/j.mex.2025.103517>
- Yudono, H., Hadi, S., Basuki, T. M., Savitri, E., Supangat, A. B., Putra, P. B., Wahyuningrum, N., Adi, R. N., Setiawan, O., Nandini, R., Cahyono, S. A., Auliyani, D., Maftukhakh, F., Nada, H., Pratiwi, D., & Hasani, M. (2025). Advancing air quality monitoring systems towards sustainable green development: Insight for metropolitan cities in Indonesia. *Environmental and Sustainability Indicators*, 26. <https://doi.org/10.1016/j.indic.2025.100649>