



Prediksi Konsentrasi CO(GT) Menggunakan Long Short-Term Memory pada Data Sensor Kualitas Udara IoT

Asep Arwan Sulaeman*, Candra Naya, Ahmad Turmudi Zy, Riyadi

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}asep.arwan@pelitabangsa.ac.id ²candranaya@pelitabangsa.ac.id, ³turmudi@pelitabangsa.ac.id, ⁴riyadi@pelitabangsa.ac.id

Email Penulis Korespondensi: asep.arwan@pelitabangsa.ac.id

Abstrak—Kualitas udara yang terus mengalami penurunan di berbagai kawasan perkotaan menjadi tantangan penting bagi kesehatan masyarakat dan pengelolaan lingkungan. Sistem pemantauan berbasis Internet of Things (IoT) menghasilkan data sensor secara kontinu yang berpotensi dimanfaatkan untuk prediksi kualitas udara, namun karakteristik data tersebut umumnya mengandung nilai hilang, derau, dan ketergantungan temporal yang dapat menurunkan kinerja model prediksi. Penelitian ini bertujuan membangun model prediksi konsentrasi karbon monoksida (CO(GT)) menggunakan algoritma Long Short-Term Memory (LSTM) pada dataset Air Quality UCI yang terdiri atas 9.357 observasi dan 15 atribut. Tahap preprocessing dilakukan dengan mengidentifikasi nilai -200 sebagai indikator *missing value*, membersihkan data yang tidak valid, melakukan normalisasi menggunakan Min-Max Scaling, serta membentuk data sekuens melalui pendekatan *sliding window* berukuran empat sebelum dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20. Model LSTM dibangun menggunakan satu lapisan LSTM berisi 50 unit dan satu lapisan Dense sebagai keluaran, kemudian dilatih menggunakan optimizer Adam selama 50 epoch. Hasil evaluasi menunjukkan nilai MAE sebesar 0,0389 dan RMSE sebesar 0,0567, yang menunjukkan kemampuan model dalam mengikuti pola temporal data kualitas udara dengan tingkat kesalahan yang relatif rendah. Hasil tersebut sejalan dengan berbagai penelitian sebelumnya yang menunjukkan bahwa LSTM efektif dalam pemodelan deret waktu kualitas udara, sehingga model yang dikembangkan berpotensi mendukung sistem pemantauan kualitas udara berbasis IoT secara berkelanjutan.

Kata kunci: Kualitas Udara; Internet of Things; Long Short-Term Memory; Prediksi; Air Quality UCI

Abstract—Air quality deterioration has become a major challenge for public health and environmental management in urban areas. Internet of Things (IoT)-based monitoring systems continuously generate sensor data that can be exploited for air quality prediction; however, these datasets commonly contain missing values, noise, and temporal dependencies that may reduce prediction accuracy. This study proposes a Long Short-Term Memory (LSTM)-based model to predict carbon monoxide (CO(GT)) concentrations using the Air Quality UCI dataset, which consists of 9,357 observations and 15 attributes. During preprocessing, -200 values were identified as missing-value indicators, followed by invalid-data handling, Min-Max normalization, and sequence generation using a sliding-window approach with a window size of four. The processed data were divided into training and testing sets using an 80:20 ratio. The prediction model employs a single LSTM layer with 50 hidden units and a Dense output layer and is trained using the Adam optimizer for 50 epochs. Experimental results achieved a Mean Absolute Error (MAE) of 0.0389 and a Root Mean Squared Error (RMSE) of 0.0567, indicating that the proposed model effectively captures temporal patterns in air quality observations with relatively low prediction errors. These findings are consistent with previous studies reporting the effectiveness of LSTM for air quality forecasting and demonstrate its potential to support continuous IoT-based environmental monitoring systems. Future work may incorporate hyperparameter optimization and comparative evaluations with alternative deep learning architectures to further improve predictive performance.

Keywords: Air Quality; Internet of Things; Long Short-Term Memory; Prediction; Air Quality UCI

1. PENDAHULUAN

Pencemaran udara merupakan salah satu penyebab utama meningkatnya risiko penyakit pernapasan, gangguan kardiovaskular, serta penurunan kualitas lingkungan di kawasan perkotaan [1], [2]. Ketersediaan sistem prediksi kualitas udara yang akurat menjadi kebutuhan penting untuk mendukung sistem peringatan dini dan pengambilan keputusan dalam pengendalian pencemaran udara [3], [4].

Perkembangan Internet of Things (IoT) telah mendorong penerapan sensor lingkungan yang mampu melakukan pemantauan kualitas udara secara kontinu dan menghasilkan data dalam jumlah besar [5], [6]. Data yang dihasilkan memiliki karakteristik deret waktu (time series), sehingga menyimpan informasi mengenai perubahan konsentrasi polutan dari waktu ke waktu [7]. Meskipun demikian, data sensor IoT umumnya menghadapi berbagai permasalahan, seperti adanya nilai hilang (*missing value*), *noise*, penciran (*outlier*), dan ketergantungan temporal yang kompleks sehingga dapat menurunkan akurasi model prediksi apabila tidak ditangani dengan baik pada tahap prapemrosesan [8]. Pada dataset Air Quality UCI, misalnya, nilai -200 digunakan sebagai penanda data yang tidak tersedia (*missing value*), sehingga memerlukan proses identifikasi dan penanganan sebelum digunakan dalam proses pemodelan [9].

Dalam implementasinya, sensor IoT dapat merekam berbagai parameter kualitas udara seperti karbon monoksida (CO), nitrogen dioksida (NO₂), sulfur dioksida (SO₂), ozon (O₃), serta partikel debu (PM2.5 dan PM10) [10], [11] dalam rentang waktu tertentu [12]. Data yang dihasilkan bersifat time series karena memiliki urutan waktu yang berkesinambungan. Karakteristik data time series memungkinkan dilakukannya analisis pola perubahan kualitas udara dari waktu ke waktu, termasuk identifikasi tren, musiman, maupun fluktuasi polutan. Namun demikian, volume data sensor IoT yang terus bertambah secara berkelanjutan menimbulkan tantangan tidak hanya dari sisi jumlah data, tetapi juga dari aspek efisiensi komputasi. Semakin besar ukuran dataset, semakin tinggi kebutuhan memori, waktu pelatihan model, dan kompleksitas proses pembelajaran, sehingga berpotensi menurunkan efisiensi sistem prediksi, terutama pada aplikasi yang memerlukan pembaruan informasi secara cepat. Oleh karena itu, dibutuhkan metode analisis yang mampu



memodelkan ketergantungan temporal secara efektif sekaligus mempertahankan efisiensi proses pembelajaran sehingga prediksi kualitas udara dapat dilakukan secara akurat dan tepat waktu.

Berbagai penelitian sebelumnya telah mengembangkan metode prediksi kualitas udara menggunakan pendekatan statistik maupun pembelajaran mesin. Metode statistik seperti Autoregressive Integrated Moving Average (ARIMA) cukup efektif untuk memodelkan pola linier, namun performanya cenderung menurun ketika menghadapi hubungan yang bersifat nonlinier dan dinamika lingkungan yang kompleks [13]. Sebaliknya, pendekatan *deep learning*, khususnya Long Short-Term Memory (LSTM), memiliki kemampuan untuk mempelajari ketergantungan jangka panjang pada data sekuensial sehingga banyak digunakan dalam prediksi kualitas udara [14]. Meskipun LSTM telah banyak diterapkan untuk prediksi kualitas udara, sebagian besar penelitian lebih menitikberatkan pada peningkatan akurasi model tanpa memberikan penjelasan yang memadai mengenai penanganan karakteristik dataset Air Quality UCI, khususnya penggunaan nilai -200 sebagai indikator missing value. Kurangnya transparansi pada tahap preprocessing menyebabkan hasil penelitian sulit direproduksi dan dibandingkan secara objektif.

Time series berbasis machine learning merupakan metode yang memiliki kemampuan dalam menganalisis pola data berdasarkan urutan waktu sehingga dapat menghasilkan prediksi yang lebih akurat terhadap kondisi kualitas udara di masa mendatang [15], [16]. Salah satu metode yang banyak digunakan adalah Long Short-Term Memory (LSTM) [17], [18], yaitu pengembangan dari Recurrent Neural Network (RNN) [19], [20] yang dirancang untuk mengatasi permasalahan vanishing gradient pada data sekuensial.

Metode LSTM memiliki kemampuan mengingat informasi jangka panjang sehingga sangat sesuai digunakan pada data sensor IoT yang bersifat kontinu dan dinamis [21], [22]. Kekuatan utama metode ini terletak pada kemampuannya menangkap hubungan temporal yang kompleks dan menghasilkan prediksi dengan tingkat akurasi tinggi. Selain itu, metode ini juga mampu mengolah data dalam jumlah besar secara lebih adaptif dibandingkan metode statistik konvensional.

Meskipun memiliki berbagai keunggulan, penggunaan metode LSTM pada dataset besar juga menghadapi beberapa tantangan. Proses pelatihan model membutuhkan waktu komputasi yang relatif lama dan sumber daya perangkat keras yang tinggi, terutama ketika data sensor IoT terus bertambah secara real-time [23]. Selain itu, kualitas data sensor yang sering mengandung missing value, noise, dan inkonsistensi dapat memengaruhi performa model prediksi. Oleh karena itu, diperlukan tahapan preprocessing data yang baik agar model mampu bekerja secara optimal. Efisiensi pengolahan data menjadi aspek penting dalam penelitian ini karena sistem prediksi kualitas udara diharapkan dapat digunakan secara berkelanjutan dan real-time pada lingkungan IoT [24].

Berdasarkan kondisi tersebut, penelitian ini mengembangkan model prediksi konsentrasi karbon monoksida (CO(GT)) menggunakan algoritma Long Short-Term Memory (LSTM) pada dataset Air Quality UCI. Tahapan penelitian diawali dengan *data cleaning*, normalisasi menggunakan Min-Max Scaling, serta pembentukan data sekuens melalui pendekatan *sliding window* sebelum proses pelatihan model.

Tahapan *preprocessing* tersebut bertujuan menghasilkan data masukan yang lebih konsisten sehingga hubungan temporal antarobservasi dapat dipelajari secara lebih efektif oleh model LSTM. Ukuran *sliding window* sebesar empat dipilih untuk merepresentasikan empat observasi historis terdekat sebagai masukan model dengan tetap menjaga kompleksitas komputasi dan panjang sekuens pada konfigurasi dasar penelitian. Meskipun demikian, ukuran *window* tersebut belum diperoleh melalui proses optimasi hiperparameter sehingga masih menjadi salah satu keterbatasan penelitian dan berpotensi ditingkatkan pada penelitian selanjutnya.

Kontribusi penelitian ini terletak pada penyusunan alur *preprocessing* yang sistematis, transparan, dan dapat direproduksi untuk pemodelan kualitas udara menggunakan dataset Air Quality UCI. Alur tersebut mengintegrasikan proses *data cleaning*, normalisasi, pembentukan data sekuens, hingga pembangunan model LSTM dalam satu tahapan yang terdokumentasi dengan jelas sehingga memudahkan reproduksi hasil penelitian. Selain itu, penelitian ini menyediakan konfigurasi dasar (*baseline*) penerapan LSTM pada prediksi konsentrasi CO(GT) yang dapat dijadikan acuan untuk pengembangan penelitian selanjutnya melalui optimasi hiperparameter maupun perbandingan dengan metode prediksi lainnya.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan prediksi deret waktu (*time series forecasting*) menggunakan algoritma Long Short-Term Memory (LSTM). Pendekatan ini dipilih karena LSTM merupakan salah satu metode *deep learning* yang dirancang untuk menangani data sekuensial serta mampu mempelajari ketergantungan jangka panjang (*long-term dependencies*) pada data deret waktu. Dengan kemampuan tersebut, model LSTM dinilai efektif dalam mengidentifikasi pola historis sehingga dapat menghasilkan prediksi yang lebih akurat dibandingkan metode konvensional pada data yang memiliki karakteristik temporal.

Tahapan penelitian diawali dengan proses pengumpulan data dari sumber yang telah ditentukan, kemudian dilanjutkan dengan tahap preprocessing untuk meningkatkan kualitas data. Proses *preprocessing* meliputi penanganan nilai yang hilang (*missing values*), normalisasi data, serta transformasi data ke dalam bentuk yang sesuai untuk model LSTM. Selanjutnya dilakukan pembentukan dataset *time series* menggunakan teknik *sliding window* agar data historis dapat digunakan sebagai masukan (*input*) untuk memprediksi nilai pada periode berikutnya. Dataset yang telah terbentuk kemudian dibagi menjadi data *training* dan data *testing* sesuai proporsi yang telah ditentukan.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan dataset Air Quality UCI yang diperoleh dari platform Kaggle. Dataset ini berisi data kualitas udara berbasis time series yang diperoleh dari sensor lingkungan, meliputi parameter polutan udara seperti CO, NO_x, NO₂, suhu, dan kelembapan. Data tersebut digunakan sebagai data utama dalam proses analisis dan prediksi kualitas udara menggunakan metode Long Short-Term Memory (LSTM).

Tabel 1. Parameter Dataset Kualitas Udara

No	Parameter	Keterangan
1	Date	Tanggal pengamatan
2	Time	Waktu pengamatan
3	CO(GT)	Konsentrasi karbon monoksida
4	PT08.S1(CO)	Sensor respon CO
5	NMHC(GT)	Konsentrasi Non-Methane Hydrocarbons
6	C6H6(GT)	Konsentrasi Benzene
7	PT08.S2(NMHC)	Sensor respon NMHC
8	NO _x (GT)	Konsentrasi Nitrogen Oxides
9	PT08.S3(NO _x)	Sensor respon NO _x
10	NO ₂ (GT)	Konsentrasi Nitrogen Dioxide
11	PT08.S4(NO ₂)	Sensor respon NO ₂
12	PT08.S5(O ₃)	Sensor respon Ozon
13	T	Suhu udara
14	RH	Kelembapan relatif
15	AH	Kelembapan absolut

Tabel 1 menunjukkan parameter-parameter yang terdapat pada dataset kualitas udara, meliputi data polutan udara, sensor pendeteksi, suhu, kelembapan, serta waktu pengamatan yang digunakan dalam proses analisis dan prediksi kualitas udara berbasis time series.

2.2. Preprocessing Data

Preprocessing data dilakukan untuk membersihkan dan menyiapkan dataset sebelum digunakan pada proses pelatihan model LSTM. Tahapan ini meliputi pembersihan data, penanganan missing value, normalisasi data, transformasi data time series, serta pembagian data menjadi training dan testing agar model dapat bekerja lebih optimal.

Tabel 2. Tahapan Preprocessing Data

No	Tahapan	Keterangan
1	Data Cleaning	Membersihkan data tidak valid dan duplikat
2	Missing Value Handling	Menangani data kosong
3	Normalisasi Data	Menyamakan skala data
4	Transformasi Time Series	Membentuk data sequence
5	Data Splitting	Membagi data training dan testing

Tabel 2 menunjukkan tahapan preprocessing data yang dilakukan untuk membersihkan, menormalkan, dan menyiapkan dataset kualitas udara sebelum digunakan pada proses pelatihan model LSTM.



2.2.1. Data Cleaning dan Missing Value Handling

Data diperiksa menggunakan fungsi `isnull()` dan pengecekan nilai -200 yang sering digunakan sebagai indikator missing value pada dataset kualitas udara. Pada tahap preprocessing dilakukan identifikasi nilai -200 sebagai indikator missing value sesuai dokumentasi Air Quality UCI. Nilai tersebut kemudian ditangani melalui proses data cleaning sehingga dataset yang digunakan untuk pemodelan tidak lagi mengandung missing value.

2.2.2. Normalisasi Data

Normalisasi data dilakukan menggunakan metode Min-Max Scaling untuk mentransformasikan seluruh atribut numerik ke dalam rentang nilai 0–1. Metode ini dipilih karena mampu mempertahankan distribusi relatif antar data sekaligus menyamakan skala setiap atribut sehingga tidak ada variabel yang mendominasi proses pembelajaran model. Selain itu, normalisasi membantu mempercepat proses konvergensi dan meningkatkan stabilitas pelatihan LSTM yang menggunakan fungsi aktivasi *sigmoid* dan *tanh*. Dengan demikian, model dapat mempelajari pola temporal pada data kualitas udara secara lebih efektif dan menghasilkan prediksi yang lebih konsisten.

2.2.3. Transformasi Data Time Series

Data target CO(GT) diubah menjadi bentuk *sequence* menggunakan pendekatan sliding window dengan ukuran *window* sebesar 4, di mana setiap empat observasi sebelumnya digunakan sebagai masukan untuk memprediksi satu observasi berikutnya. Pendekatan ini bertujuan mengubah data deret waktu menjadi sekuens yang dapat dipelajari oleh model Long Short-Term Memory (LSTM). Pemilihan *window size* sebesar empat dilakukan untuk merepresentasikan hubungan temporal jangka pendek sekaligus menjaga kompleksitas model agar tetap rendah pada penelitian ini. Meskipun demikian, ukuran *window* tersebut belum diperoleh melalui proses optimasi hiperparameter sehingga masih menjadi salah satu keterbatasan penelitian. Penelitian selanjutnya dapat mengevaluasi berbagai ukuran *window* yang lebih panjang untuk memperoleh konfigurasi yang memberikan performa prediksi terbaik.

2.2.4. Data Splitting

Setelah proses pembentukan data sekuens selesai, dataset dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan rasio 80:20. Berbeda dengan pembagian data pada kasus klasifikasi atau regresi umum, proses pembagian pada penelitian ini dilakukan secara kronologis tanpa proses pengacakan (*shuffle=False*), sehingga urutan temporal pada data tetap dipertahankan. Sebanyak 80% observasi awal digunakan sebagai data pelatihan, sedangkan 20% observasi terakhir digunakan sebagai data pengujian.

Pendekatan ini dipilih karena data kualitas udara merupakan data *time series* yang memiliki ketergantungan antarwaktu. Pengacakan data berpotensi menyebabkan data leakage, yaitu kondisi ketika informasi dari periode mendatang secara tidak langsung digunakan pada proses pelatihan sehingga menghasilkan estimasi performa model yang terlalu optimistis. Dengan mempertahankan urutan kronologis, proses evaluasi menjadi lebih realistis karena model hanya memanfaatkan informasi historis untuk memprediksi observasi pada periode berikutnya.

2.3. Pembentukan Dataset Time Series

Tahap pembentukan dataset time series dilakukan dengan menyusun data kualitas udara berdasarkan urutan waktu pengamatan agar dapat diproses oleh model Long Short-Term Memory (LSTM). Pada tahap ini, data diubah menjadi bentuk *sequence* atau deret waktu sehingga model dapat mempelajari pola perubahan kualitas udara dari waktu sebelumnya untuk memprediksi kondisi pada waktu berikutnya.

Tabel 3. Konfigurasi Pembentukan Dataset Time Series

Parameter	Nilai
Metode	Sliding Window
Target Prediksi	CO(GT)
Window Size	4
Jumlah Record Awal	9.357
Jumlah Sequence	9.353
Jumlah Target	9.353

Tabel 3 menunjukkan konfigurasi yang digunakan dalam proses pembentukan dataset time series. Metode yang digunakan adalah *sliding window* dengan ukuran jendela (*window size*) sebesar 4. Variabel target yang diprediksi adalah konsentrasi karbon monoksida CO(GT). Empat data pengamatan sebelumnya digunakan sebagai variabel input untuk memprediksi satu nilai CO(GT) pada periode berikutnya. Konfigurasi tersebut dipilih agar model LSTM dapat mempelajari pola temporal dan hubungan antar data secara berurutan sesuai karakteristik data time series kualitas udara.

Tabel 4. Contoh Pembentukan Sequence Data

Sequence	X1	X2	X3	X4	Target (y)
1	0,211864	0,161017	0,177966	0,177966	0,127119
2	0,161017	0,177966	0,177966	0,127119	0,093220



Tabel 4 memperlihatkan contoh hasil transformasi data kualitas udara ke dalam bentuk sequence yang digunakan sebagai input model LSTM. Setiap sequence terdiri atas empat nilai historis yang berperan sebagai variabel input (X1–X4), sedangkan nilai berikutnya digunakan sebagai target prediksi (y). Sebagai contoh, sequence pertama menggunakan empat nilai data sebelumnya, yaitu 0,211864; 0,161017; 0,177966; dan 0,177966 untuk memprediksi nilai target sebesar 0,127119.

2.4. Pembagian Data Training dan Testing

Tahap pembagian data dilakukan untuk memisahkan dataset menjadi data training dan data testing. Data training digunakan untuk melatih model Long Short-Term Memory (LSTM) dalam mempelajari pola kualitas udara, sedangkan data testing digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya.

2.5. Perancangan Model LSTM

Tahap perancangan model dilakukan dengan membangun arsitektur LSTM dan menentukan parameter pelatihan seperti epoch, batch size, dan optimizer untuk memproses data kualitas udara berbasis time series.

Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Cell State

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3)$$

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

Persamaan tersebut digunakan untuk mengatur proses penyimpanan dan pembaruan informasi pada model LSTM sehingga model mampu menghasilkan prediksi kualitas udara yang lebih akurat berdasarkan pola data time series.

2.6. Pelatihan Model

Tahap pelatihan dilakukan menggunakan data training untuk mempelajari pola kualitas udara berdasarkan data historis. Proses pelatihan menggunakan fungsi loss Mean Squared Error (MSE) guna mengukur kesalahan prediksi model.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

n = jumlah data

Nilai *loss* digunakan untuk mengukur tingkat kesalahan prediksi selama proses pelatihan. Semakin kecil nilai *loss*, maka semakin baik kemampuan model dalam mempelajari pola data kualitas udara.

2.7. Pengujian Model

Model Long Short-Term Memory (LSTM) dilatih menggunakan data pelatihan yang telah melalui tahap normalisasi dan pembentukan data sekuens. Proses pelatihan bertujuan mempelajari pola temporal pada data historis sehingga model mampu memprediksi nilai konsentrasi CO(GT) pada periode berikutnya. Pada penelitian ini digunakan optimizer Adam karena memiliki kemampuan konvergensi yang baik pada permasalahan *deep learning*, dengan learning rate sebesar 0,001, batch size 32, dan 50 epoch. Arsitektur model terdiri atas satu lapisan LSTM dengan 50 unit neuron yang diikuti oleh satu lapisan Dense sebagai keluaran. Konfigurasi hiperparameter tersebut digunakan sebagai konfigurasi dasar (*baseline*) dan belum melalui proses optimasi, sehingga masih terbuka peluang untuk peningkatan performa.

Tabel 5. Konfigurasi Hiperparameter Model LSTM

Parameter	Nilai
Arsitektur	1 LSTM + 1 Dense
Jumlah Unit LSTM	50
Optimizer	Adam
Learning Rate	0,001
Batch Size	32
Epoch	50
Loss Function	Mean Squared Error (MSE)



Parameter	Nilai
Aktivasi LSTM	tanh
Aktivasi Output	Linear

2.8. Evaluasi Model

Tahap evaluasi dilakukan dengan membandingkan hasil prediksi dan data aktual menggunakan metrik Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) untuk mengukur kinerja model LSTM.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Keterangan:

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

n = jumlah data

Nilai MAE dan RMSE digunakan untuk menentukan tingkat akurasi model. Model dengan nilai error yang lebih kecil menunjukkan kemampuan prediksi yang lebih baik terhadap data kualitas udara.

3. HASIL DAN PEMBAHASAN

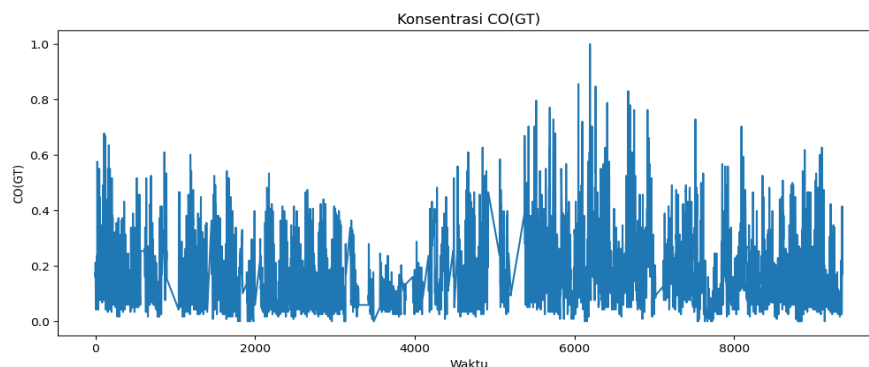
3.1. Hasil Preprocessing Data

Tahap preprocessing meliputi data cleaning, normalisasi menggunakan Min-Max Scaling, pembentukan data time series, serta pembagian data menjadi training dan testing. Hasilnya menunjukkan bahwa dataset telah bersih dan siap digunakan untuk pelatihan model LSTM.

Tabel 6. Dataset Setelah Preprocessing

Parameter	Nilai
Jumlah Record	9.357
Jumlah Atribut	15
Missing Value	0
Data Duplikat	0
Metode Normalisasi	Min-Max Scaling
Rentang Data	0–1
Status Dataset	Siap untuk pelatihan model LSTM

Tabel 6 menunjukkan hasil preprocessing dataset kualitas udara yang terdiri atas 9.357 record dan 15 atribut. Hasil pemeriksaan menunjukkan tidak terdapat missing value maupun data duplikat sehingga dataset layak digunakan untuk pemodelan. Selanjutnya dilakukan normalisasi menggunakan metode Min-Max Scaling dengan rentang data 0–1 untuk meningkatkan stabilitas dan kinerja model LSTM selama proses pelatihan.



Gambar 2. Visualisasi Data CO(GT) Setelah Normalisasi

Gambar 2 menunjukkan hasil visualisasi data target CO(GT) setelah dilakukan proses normalisasi menggunakan metode Min-Max Scaling. Seluruh nilai konsentrasi CO(GT) berhasil ditransformasikan ke dalam rentang 0–1 tanpa mengubah pola distribusi maupun karakteristik temporal data. Terlihat bahwa fluktuasi nilai CO(GT) tetap dipertahankan

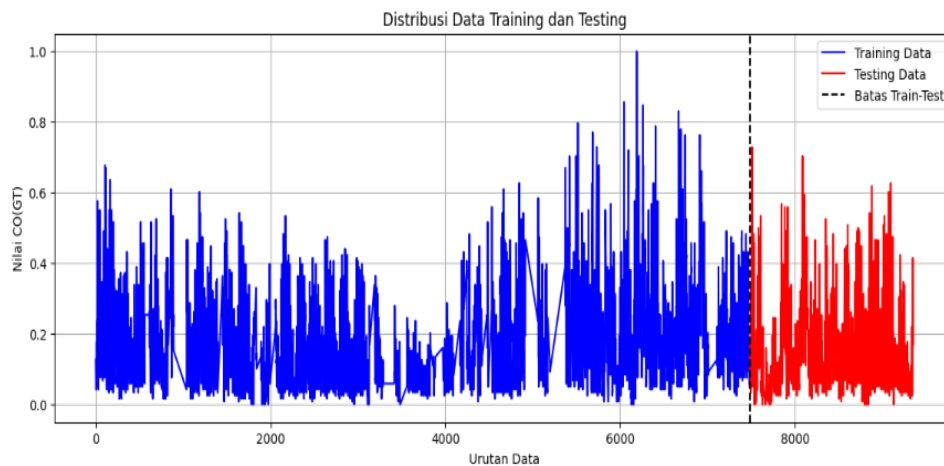
setelah normalisasi, sehingga informasi mengenai kenaikan dan penurunan konsentrasi polutan tidak mengalami distorsi. Kondisi ini menunjukkan bahwa proses normalisasi berhasil menyamakan skala data sekaligus mempertahankan pola historis yang diperlukan oleh model Long Short-Term Memory (LSTM) dalam mempelajari hubungan temporal antarobservasi. Dengan demikian, data hasil normalisasi menjadi lebih sesuai sebagai masukan model, membantu mempercepat proses konvergensi selama pelatihan, serta mengurangi pengaruh perbedaan skala terhadap performa prediksi

3.2. Hasil Pembentukan Dataset Time Series

Setelah melalui tahap preprocessing, data kualitas udara disusun berdasarkan urutan waktu pengamatan untuk membentuk dataset *time series*. Proses ini dilakukan agar model Long Short-Term Memory (LSTM) dapat mempelajari pola perubahan kualitas udara secara berurutan dari waktu ke waktu. Pada penelitian ini, data yang telah dibersihkan dan dinormalisasi kemudian diurutkan secara kronologis sebelum diubah menjadi bentuk *sequence*.

3.3. Hasil Pembagian Data Training dan Testing

Setelah data berhasil dibentuk menjadi dataset *time series*, tahap berikutnya adalah pembagian data menjadi data *training* dan *testing*. Pembagian data dilakukan untuk memastikan model Long Short-Term Memory (LSTM) dapat dilatih menggunakan sebagian data dan diuji menggunakan data yang belum pernah dipelajari sebelumnya. Pada penelitian ini digunakan rasio pembagian data sebesar **80:20**, di mana 80% data digunakan sebagai data *training* dan 20% sisanya digunakan sebagai data *testing*.



Gambar 3. Distribusi Data Training dan Testing

Gambar 3 menunjukkan hasil pembagian dataset menjadi data training dan data testing menggunakan rasio 80:20. Data training ditunjukkan oleh garis berwarna biru yang digunakan untuk proses pelatihan model LSTM, sedangkan data testing ditunjukkan oleh garis berwarna merah yang digunakan untuk mengevaluasi kemampuan model. Garis putus-putus menunjukkan batas antara data training dan data testing. Hasil visualisasi menunjukkan bahwa distribusi data testing masih memiliki pola yang serupa dengan data training sehingga layak digunakan untuk proses evaluasi model.

3.4. Hasil Perancangan Model LSTM

Setelah dataset dibagi menjadi data *training* dan *testing*, dilakukan perancangan model Long Short-Term Memory (LSTM) untuk mempelajari pola temporal pada data kualitas udara. Model dirancang menggunakan satu lapisan LSTM dan satu lapisan output untuk menghasilkan nilai prediksi kualitas udara. Konfigurasi model yang digunakan ditunjukkan pada Tabel 6.

Tabel 7. Arsitektur Model LSTM Hasil Implementasi

Layer	Output Shape	Parameter
LSTM	(None, 50)	10.400
Dense	(None, 1)	51
Total Trainable Parameters	-	10.451

Tabel 7 menunjukkan arsitektur model LSTM yang digunakan dalam penelitian. Model terdiri atas satu lapisan LSTM dengan 50 unit dan satu lapisan Dense sebagai output prediksi. Total parameter yang dilatih sebanyak 10.451 parameter untuk mempelajari pola data kualitas udara dan menghasilkan prediksi nilai CO(GT).

3.5. Hasil Pelatihan Model

Tahap pelatihan model dilakukan menggunakan data training yang telah melalui proses preprocessing, pembentukan data *time series*, dan pembagian data menjadi data pelatihan serta data pengujian. Pada tahap ini, model Long Short-Term

Memory (LSTM) dilatih untuk mempelajari pola perubahan kualitas udara berdasarkan data historis yang diperoleh dari sensor IoT. Data yang telah dinormalisasi dan disusun dalam bentuk sequence digunakan sebagai input agar model mampu mengenali hubungan antarwaktu dan menangkap pola perubahan kualitas udara secara lebih efektif.

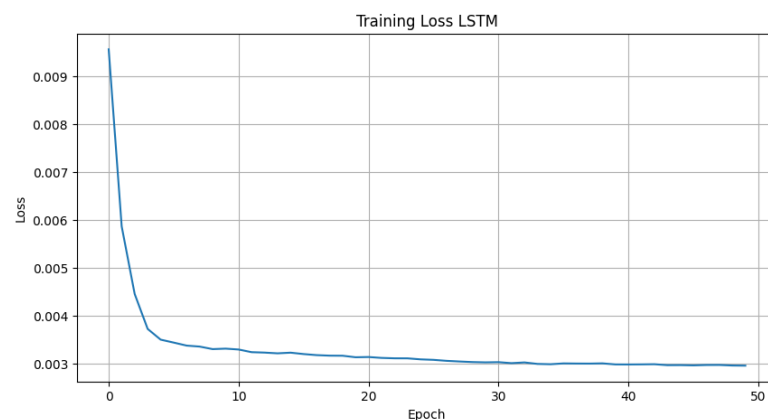
Model LSTM dikompilasi menggunakan optimizer Adam dan fungsi loss Mean Squared Error (MSE). Optimizer Adam dipilih karena memiliki kemampuan untuk menyesuaikan laju pembelajaran secara adaptif sehingga dapat mempercepat proses konvergensi model. Pelatihan dilakukan selama 50 epoch, di mana pada setiap epoch model melakukan proses prediksi terhadap data training, menghitung nilai error, dan memperbarui bobot jaringan melalui mekanisme backpropagation. Proses ini berlangsung secara berulang hingga model memperoleh parameter yang optimal untuk menghasilkan prediksi yang akurat.

Selama proses pelatihan, nilai loss diamati pada setiap epoch untuk mengevaluasi kemampuan model dalam mempelajari pola data. Nilai loss yang semakin kecil menunjukkan bahwa selisih antara hasil prediksi dan data aktual semakin rendah, sehingga menandakan peningkatan kinerja model. Pemantauan nilai loss juga penting untuk memastikan proses pelatihan berjalan dengan baik dan model tidak mengalami kesulitan dalam mempelajari pola data. Setelah proses pelatihan selesai, model yang telah terbentuk digunakan untuk melakukan prediksi kualitas udara pada data pengujian dan dievaluasi menggunakan metrik performa yang sesuai.

Tabel 7. Hasil Pelatihan Model LSTM

Epoch	Loss
1	0,0095
10	0,0033
20	0,0031
30	0,0030
40	0,0030
50	0,0030

Tabel 7 menunjukkan perkembangan nilai *loss* selama proses pelatihan model LSTM. Terlihat bahwa nilai *loss* mengalami penurunan yang signifikan dari 0,0095 pada *epoch* pertama menjadi 0,0030 pada *epoch* ke-30 dan tetap stabil hingga *epoch* ke-50. Hal ini mengindikasikan bahwa model telah mencapai kondisi konvergen dan mampu mempelajari pola temporal pada data pelatihan dengan baik.



Gambar 4. Grafik Perkembangan Nilai Loss Selama Proses Pelatihan Model LSTM

Gambar 4 memperlihatkan perubahan nilai *training loss* selama proses pelatihan model LSTM selama 50 *epoch*. Terlihat bahwa nilai *loss* menurun secara signifikan pada beberapa *epoch* awal, kemudian berangsur menurun hingga mencapai kondisi yang stabil pada kisaran **0,0030** setelah sekitar *epoch* ke-30. Pola tersebut menunjukkan bahwa proses optimasi berjalan dengan baik dan model berhasil mencapai konvergensi. Meskipun demikian, karena penelitian ini hanya menampilkan *training loss* tanpa *validation loss*, evaluasi terhadap potensi *overfitting* belum dapat dilakukan secara menyeluruh.

3.6. Hasil Pengujian Model

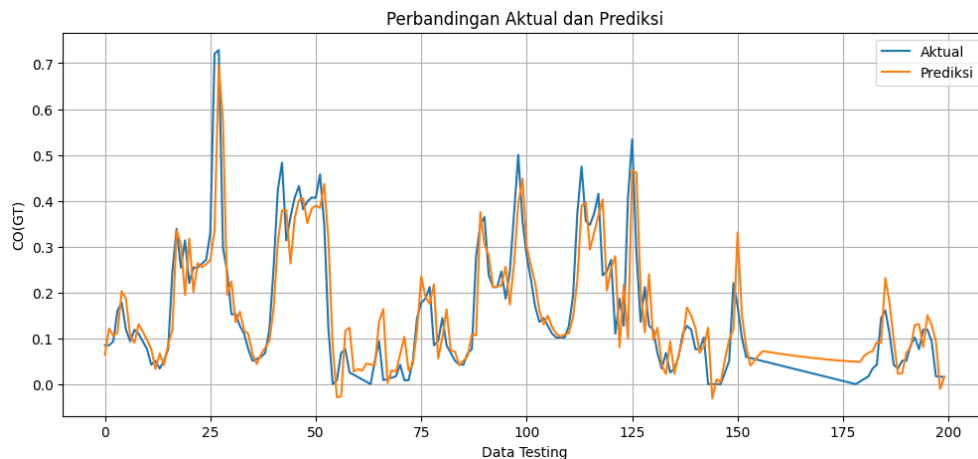
Setelah proses pelatihan selesai, model Long Short-Term Memory (LSTM) diuji menggunakan data testing yang tidak terlibat dalam proses pembelajaran. Pengujian ini bertujuan untuk mengetahui kemampuan model dalam melakukan prediksi kualitas udara pada data yang belum pernah dilihat sebelumnya. Dengan menggunakan data testing, dapat diketahui sejauh mana model mampu menggeneralisasi pola yang telah dipelajari dari data training ke data baru.

Hasil prediksi yang dihasilkan model kemudian dibandingkan dengan nilai aktual untuk menilai tingkat akurasi dan kinerja model. Evaluasi dilakukan menggunakan beberapa metrik, seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2). Selain itu, perbandingan antara nilai prediksi dan nilai aktual juga divisualisasikan dalam bentuk grafik untuk memberikan gambaran yang lebih jelas mengenai kemampuan model dalam mengikuti pola perubahan kualitas udara dari waktu ke waktu.

Tabel 9. Hasil Prediksi pada Data Testing

No	Nilai Aktual (y_{test})	Nilai Prediksi (y_{pred})	Selisih Absolut
1	0,064518	0,187188	0,122670
2	0,121254	0,200583	0,079329
3	0,165286	0,165392	0,000105

Tabel 9 menunjukkan contoh hasil prediksi model LSTM pada data testing. Secara umum, nilai prediksi yang dihasilkan memiliki pola yang mendekati nilai aktual meskipun masih terdapat selisih pada beberapa data. Hasil ini menunjukkan bahwa model LSTM mampu mempelajari pola data kualitas udara dengan baik dan menghasilkan prediksi yang cukup akurat terhadap data yang belum pernah digunakan pada proses pelatihan.

**Gambar 5.** Perbandingan Nilai Aktual dan Prediksi Menggunakan LSTM

Grafik pada Gambar 5 menunjukkan bahwa kurva hasil prediksi memiliki pola yang sangat dekat dengan kurva data aktual. Kedekatan kedua kurva tersebut mengindikasikan bahwa model LSTM mampu memprediksi kualitas udara dengan tingkat kesalahan yang relatif kecil.

Berdasarkan hasil pengujian, model LSTM berhasil menghasilkan prediksi yang mendekati nilai aktual pada sebagian besar data *testing*. Selisih prediksi yang relatif kecil menunjukkan bahwa model mampu menangkap pola temporal dalam data kualitas udara dengan baik. Hasil pengujian ini menjadi dasar untuk tahap evaluasi model menggunakan metrik MAE dan RMSE pada subbab berikutnya.

3.7. Evaluasi Kinerja Model

Tahap evaluasi dilakukan untuk mengukur kinerja model Long Short-Term Memory (LSTM) dalam memprediksi kualitas udara. Evaluasi dilakukan menggunakan metrik Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) dengan membandingkan nilai prediksi terhadap nilai aktual. Semakin kecil nilai error yang dihasilkan, maka semakin baik performa model dalam melakukan prediksi.

Tabel 10. Hasil Evaluasi Kinerja Model

No	Metrik Evaluasi	Nilai
1	Mean Absolute Error (MAE)	0,0389
2	Root Mean Squared Error (RMSE)	0,0567

Tabel 10 menunjukkan hasil evaluasi model LSTM berdasarkan data pengujian. Nilai MAE sebesar 0,0389 menunjukkan bahwa rata-rata kesalahan prediksi model relatif kecil, sedangkan nilai RMSE sebesar 0,0567 menunjukkan bahwa penyimpangan antara nilai aktual dan nilai prediksi masih berada pada tingkat yang rendah. Nilai MAE sebesar 0,0389 pada data yang telah dinormalisasi menunjukkan bahwa rata-rata penyimpangan prediksi relatif kecil terhadap rentang data yang digunakan. Hal ini mengindikasikan bahwa model mampu mempertahankan pola temporal utama konsentrasi CO(GT). Namun demikian, karena evaluasi dilakukan pada data hasil normalisasi, interpretasi signifikansi praktis kesalahan prediksi dalam satuan asli konsentrasi CO(GT) masih terbatas. Oleh sebab itu, penelitian selanjutnya perlu menyajikan hasil evaluasi setelah proses *inverse transformation* agar tingkat kesalahan prediksi dapat diinterpretasikan secara lebih langsung.

Berdasarkan hasil evaluasi tersebut, model LSTM menunjukkan kemampuan yang baik dalam mempelajari pola temporal pada dataset Air Quality UCI yang digunakan dalam penelitian ini. Namun demikian, penelitian ini belum melakukan perbandingan dengan metode prediksi lain, seperti ARIMA, Simple RNN, maupun GRU, sehingga hasil yang diperoleh belum dapat digunakan untuk menyimpulkan bahwa LSTM merupakan metode terbaik. Oleh karena itu, penelitian selanjutnya perlu melakukan evaluasi komparatif terhadap beberapa algoritma prediksi agar keunggulan relatif model LSTM dapat dibuktikan secara lebih objektif.



3.8. Pembahasan

Hasil penelitian menunjukkan bahwa model Long Short-Term Memory (LSTM) mampu mempelajari pola temporal pada data kualitas udara dengan baik, yang ditunjukkan oleh nilai MAE sebesar **0,0389** dan RMSE sebesar **0,0567**. Nilai kesalahan yang relatif rendah mengindikasikan bahwa model berhasil menangkap hubungan antarobservasi pada data deret waktu sehingga mampu menghasilkan prediksi yang mendekati nilai aktual. Kemampuan tersebut didukung oleh karakteristik LSTM yang dirancang untuk mempertahankan informasi jangka pendek maupun jangka panjang melalui mekanisme *memory cell* dan *gating*, sehingga lebih sesuai untuk memodelkan data kualitas udara dibandingkan metode jaringan saraf konvensional.

Hasil penelitian ini juga sejalan dengan berbagai penelitian sebelumnya yang melaporkan bahwa LSTM memiliki kemampuan yang baik dalam memodelkan data kualitas udara berbasis *time series*. Dibandingkan dengan metode statistik seperti ARIMA, LSTM memiliki keunggulan dalam mempelajari hubungan nonlinier antarvariabel tanpa memerlukan asumsi stasioneritas data. Selain itu, beberapa penelitian terdahulu juga menunjukkan bahwa LSTM mampu memberikan performa yang lebih baik dibandingkan Simple RNN karena mekanisme *forget gate* dan *input gate* dapat mengurangi permasalahan *vanishing gradient* pada data sekuens yang panjang.

Meskipun demikian, penelitian ini belum melakukan evaluasi komparatif terhadap metode lain, seperti ARIMA, GRU, maupun Simple RNN. Oleh karena itu, hasil yang diperoleh belum dapat digunakan untuk menyimpulkan bahwa LSTM merupakan metode terbaik untuk prediksi kualitas udara. Penelitian ini lebih berfungsi sebagai evaluasi kemampuan model LSTM pada dataset Air Quality UCI dengan konfigurasi dasar (*baseline*). Penelitian selanjutnya perlu melakukan optimasi hiperparameter serta membandingkan beberapa algoritma prediksi agar keunggulan relatif masing-masing metode dapat dianalisis secara lebih komprehensif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Long Short-Term Memory (LSTM) memiliki kemampuan yang baik dalam memodelkan pola temporal pada data kualitas udara setelah melalui tahapan *preprocessing* yang disesuaikan dengan karakteristik dataset Air Quality UCI, meliputi identifikasi dan penanganan nilai *missing value*, normalisasi menggunakan Min-Max Scaling, serta pembentukan data sekuens melalui pendekatan *sliding window*. Hasil penelitian menunjukkan bahwa kombinasi tahapan prapemrosesan yang tepat dengan arsitektur LSTM mampu menghasilkan model prediksi yang stabil dalam mempelajari hubungan temporal antarobservasi, sehingga dapat dimanfaatkan sebagai pendekatan yang menjanjikan untuk prediksi konsentrasi karbon monoksida (CO(GT)). Temuan ini menegaskan bahwa kualitas data masukan dan proses prapemrosesan memiliki peran yang sama pentingnya dengan pemilihan algoritma dalam meningkatkan performa model prediksi berbasis *deep learning*. Dari sisi implementasi, model yang dikembangkan berpotensi diintegrasikan ke dalam sistem pemantauan kualitas udara berbasis Internet of Things (IoT) sebagai komponen prediksi dini (*early prediction*) untuk mendukung pemantauan kondisi lingkungan secara berkelanjutan, penyampaian peringatan dini terhadap peningkatan konsentrasi polutan, serta penyediaan informasi yang dapat mendukung pengambilan keputusan oleh pemerintah daerah dan instansi terkait dalam penyusunan strategi pengendalian pencemaran udara berbasis data. Meskipun demikian, penelitian ini masih menggunakan konfigurasi model dasar tanpa optimasi hiperparameter dan belum melakukan evaluasi komparatif dengan metode prediksi lain, sehingga penelitian selanjutnya disarankan melakukan optimasi parameter model, membandingkan performa LSTM dengan algoritma lain seperti GRU, Simple RNN, maupun ARIMA, serta melakukan pengujian pada dataset dari lokasi yang berbeda untuk meningkatkan kemampuan generalisasi dan keandalan model pada berbagai kondisi kualitas udara.

REFERENCES

- [1] B. Karimi and S. Samadi, "Long-term exposure to air pollution on cardio-respiratory, and lung cancer mortality: a systematic review and meta-analysis," *J. Environ. Health Sci. Eng.*, vol. 22, no. 1, pp. 75–95, May 2024, doi: 10.1007/s40201-024-00900-6.
- [2] M. S. Tavares, C. T. Rodrigues, S. L. S. de Menezes, and A. M. Moreno, "Health at Risk: Air Pollution and Urban Vulnerability—Perspectives in Light of the 2030 Agenda," *Green Health*, vol. 1, no. 3, p. 21, Nov. 2025, doi: 10.3390/greenhealth1030021.
- [3] S. Chadalavada *et al.*, "Application of artificial intelligence in air pollution monitoring and forecasting: A systematic review," *Environmental Modelling & Software*, vol. 185, p. 106312, Feb. 2025, doi: 10.1016/j.envsoft.2024.106312.
- [4] A. Garcia, Y. Saez, I. Harris, X. Huang, and E. Collado, "Advancements in air quality monitoring: a systematic review of IoT-based air quality monitoring and AI technologies," *Artif. Intell. Rev.*, vol. 58, no. 9, p. 275, Jun. 2025, doi: 10.1007/s10462-025-11277-9.
- [5] D. N. Azizah, S. Heranurweni, and L. O. M. Idris, "Internet of Things Based Air Quality Monitoring System with Automatic Notification," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 776–787, Jun. 2025, doi: 10.57152/malcom.v5i3.1945.
- [6] D. N. Azizah, S. Heranurweni, and L. O. M. Idris, "Internet of Things Based Air Quality Monitoring System with Automatic Notification," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 776–787, Jun. 2025, doi: 10.57152/malcom.v5i3.1945.
- [7] C. Liu *et al.*, "Characteristics and mechanism of a persistent ozone pollution event in Pearl River Delta induced by typhoon and subtropical high," *Atmos. Environ.*, vol. 310, p. 119964, Oct. 2023, doi: 10.1016/j.atmosenv.2023.119964.



- [8] A. E. Edje, M. S. Abd Latiff, and W. H. Chan, "IoT data analytic algorithms on edge-cloud infrastructure: A review," *Digital Communications and Networks*, vol. 9, no. 6, pp. 1486–1515, Dec. 2023, doi: 10.1016/j.dcan.2023.10.002.
- [9] F. Ares-Robledo, H. Rifà-Pous, and R. Clarisó, "Graph neural networks for anomaly detection: a systematic review of dynamic temporal approaches," *Artif. Intell. Rev.*, vol. 59, no. 5, p. 129, Mar. 2026, doi: 10.1007/s10462-026-11532-7.
- [10] D. C. Joshi *et al.*, "Fine particulate matter (PM_{2.5}, PM₁₀): A silent catalyst for chronic lung diseases in India; a comprehensive review," *Environmental Challenges*, vol. 20, p. 101215, Sep. 2025, doi: 10.1016/j.envc.2025.101215.
- [11] J. W. Sitanggang, E. Sunarsih, and H. Hasyim, "Literature Review: Analysis of Exposure of Vehicle Emission Gases (Co, No₂, So₂, Pm_{2.5}, and Pm₁₀) to Public Health Risks," *Journal of Social Research*, vol. 2, no. 7, pp. 2278–2287, Jun. 2023, doi: 10.55324/josr.v2i7.1142.
- [12] D. P. S. Sudibya, D. A. A. Afaaf, L. Satria, N. U. P. Prasna, and U. Kamal, "Implementasi Kebijakan Baku Mutu Udara Sebagai Bentuk Penanggulangan Pemanasan Global," *Jurnal Ilmu Sosial dan Humaniora*, vol. 3, no. 4, pp. 571–598, Jun. 2025, doi: 10.57248/jishum.v3i4.593.
- [13] W. M. S. Bandara and W. A. R. De Mel, "Forecasting Nonlinear Time Series with ARIMA, ANN, and Hybrid Models: A Case Study on Inflation Rate in Sri Lanka," *Indonesian Journal of Statistics and Its Applications*, vol. 9, no. 1, pp. 145–156, Jun. 2025, doi: 10.29244/ijsa.v9i1p145-156.
- [14] M. Karmoude *et al.*, "Machine learning for air quality prediction and data analysis: Review on recent advancements, challenges, and outlooks," *Science of The Total Environment*, vol. 1002, p. 180593, Nov. 2025, doi: 10.1016/j.scitotenv.2025.180593.
- [15] R. Espinosa, J. Palma, F. Jiménez, J. Kamińska, G. Sciavico, and E. Lucena-Sánchez, "A time series forecasting based multi-criteria methodology for air quality prediction," *Appl. Soft Comput.*, vol. 113, p. 107850, Dec. 2021, doi: 10.1016/j.asoc.2021.107850.
- [16] M. Rajesh, R. G. Babu, U. Moorthy, and S. V. Easwaramoorthy, "Machine learning-driven framework for realtime air quality assessment and predictive environmental health risk mapping," *Sci. Rep.*, vol. 15, no. 1, p. 28801, Aug. 2025, doi: 10.1038/s41598-025-14214-6.
- [17] Vella Puspitasari Wijayanti and Adie Wahyudi Oktavia Gama, "Analisis Model Long Short-Term Memory (LSTM) untuk Prediksi Kadar Particulate Matter 2.5," *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 6, no. 1, pp. 616–629, Mar. 2026, doi: 10.55606/juisik.v6i1.2228.
- [18] C. W. Chen Wang, B. L. Chen Wang, J. C. Bingchun Liu, and X. Y. Jiali Chen, "Air Quality Index Prediction Based on a Long Short-Term Memory Artificial Neural Network Model," *J. Comput. (Taipei)*, vol. 34, no. 2, pp. 069–079, Apr. 2023, doi: 10.53106/199115992023043402006.
- [19] S. Raheja and S. Malik, "Prediction of Air Quality Using LSTM Recurrent Neural Network," *International Journal of Software Innovation*, vol. 10, no. 1, pp. 1–16, Apr. 2022, doi: 10.4018/IJSI.297982.
- [20] Y. Sudriani, I. Ridwansyah, and H. A. Rustini, "Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 299, p. 012037, Jul. 2019, doi: 10.1088/1755-1315/299/1/012037.
- [21] R. Sudha, A. Damodaran, and G. Manohar, "Enhanced air quality prediction using adaptive residual Bi-LSTM with pyramid dilation and optimal weighted feature selection," *Sci. Rep.*, vol. 15, no. 1, p. 30428, Aug. 2025, doi: 10.1038/s41598-025-14668-8.
- [22] N. S. Muruganandam and U. Arumugam, "Seminal Stacked Long Short-Term Memory (SS-LSTM) Model for Forecasting Particulate Matter (PM_{2.5} and PM₁₀)," *Atmosphere (Basel)*, vol. 13, no. 10, p. 1726, Oct. 2022, doi: 10.3390/atmos13101726.
- [23] Dani Rofianto *et al.*, "IoT-Based Smart Irrigation and Fertilization System with Realtime Cloud Integration," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 15, no. 1, pp. 34–43, Feb. 2026, doi: 10.22146/jnteti.v15i1.23124.
- [24] M. Ansari and M. Alam, "An Intelligent IoT-Cloud-Based Air Pollution Forecasting Model Using Univariate Time-Series Analysis," *Arab. J. Sci. Eng.*, vol. 49, no. 3, pp. 3135–3162, Mar. 2024, doi: 10.1007/s13369-023-07876-9.