

# Peramalan Data Kualitas Udara Menggunakan Multivariat LSTM di Wilayah Kota Surabaya

1<sup>st</sup> Inge Faradila Efaranti  
Sains Data  
Universitas Telkom  
Surabaya, Indonesia

[ingefe@student.telkomuniversity.ac.id](mailto:ingefe@student.telkomuniversity.ac.id)

2<sup>nd</sup> Regita Putri Permata  
Sains Data  
Universitas Telkom  
Surabaya, Indonesia

[regitapermata@telkomuniversity.ac.id](mailto:regitapermata@telkomuniversity.ac.id)

3<sup>rd</sup> Rifdatun Ni'mah  
Sains Data  
Universitas Telkom  
Surabaya, Indonesia

[rifdatun@tekomuniversity.ac.id](mailto:rifdatun@tekomuniversity.ac.id)

**Abstrak** — Peningkatan polusi udara di wilayah perkotaan, termasuk Kota Surabaya, mendorong perlunya pengembangan model peramalan kualitas udara yang akurat dan adaptif. Penelitian ini bertujuan untuk menganalisis hubungan antara parameter meteorologi dan kualitas udara, serta membangun model peramalan menggunakan metode Long Short-Term Memory (LSTM) berbasis data multivariat. Data yang digunakan diperoleh dari Stasiun Pemantauan Kualitas Udara (SPKU) Kebonsari periode Januari 2022– Desember 2024, dengan parameter suhu udara, kelembapan, dan kecepatan angin sebagai input, serta PM10 dan CO sebagai target. Analisis korelasi dilakukan untuk mengidentifikasi pengaruh antar parameter, dan hasilnya menunjukkan hubungan signifikan yang dapat dimanfaatkan dalam peramalan. Model LSTM dibangun dengan pendekatan time series dan dilatih menggunakan arsitektur jaringan yang mampu menangkap pola temporal antar variabel. Evaluasi kinerja model dilakukan dengan metrik Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), dan Symmetric Mean Absolute Percentage Error (SMAPE). Hasil evaluasi menunjukkan bahwa model menghasilkan MSE sebesar 113.6211, RMSE sebesar 10.6593, MAPE sebesar 49.45%, dan SMAPE sebesar 28.17%, yang mengindikasikan performa peramalan yang cukup baik. Dengan hasil tersebut, model multivariat LSTM memiliki potensi untuk digunakan sebagai alat bantu dalam pemantauan dan pengendalian kualitas udara oleh instansi terkait di Kota Surabaya.

**Kata kunci**— CO, Kualitas Udara, LSTM, Multivariat, PM10, Peramalan

## I. PENDAHULUAN

Kualitas udara merupakan isu krusial yang semakin mendapat perhatian di berbagai kota besar, termasuk Kota Surabaya[1]. Pertumbuhan jumlah kendaraan bermotor, aktivitas industri, dan urbanisasi yang pesat telah menyebabkan peningkatan emisi polutan udara seperti partikulat (PM10) dan karbon monoksida (CO)[2]. Paparan terhadap polutan-polutan ini tidak hanya berdampak buruk terhadap kesehatan manusia[3], terutama sistem pernapasan dan kardiovaskular, tetapi juga memengaruhi ekosistem dan kualitas hidup secara umum. Oleh karena itu, pengendalian dan pemantauan kualitas udara menjadi kebutuhan yang mendesak, terutama di wilayah urban yang padat aktivitas[4].

Upaya pemerintah dalam mengawasi kualitas udara telah diwujudkan melalui pembangunan Stasiun Pemantauan Kualitas Udara (SPKU) di berbagai lokasi, salah satunya adalah SPKU Kebonsari yang berada di kawasan strategis Kota Surabaya. SPKU ini secara berkala mencatat data

lingkungan yang meliputi parameter PM10, CO, suhu udara, kelembapan, dan kecepatan angin[2]. Namun, meskipun sistem ini mampu memberikan informasi secara real-time, fungsinya masih sebatas monitoring dan belum dilengkapi dengan sistem prediktif yang dapat digunakan untuk peringatan dini atau pengambilan keputusan berbasis data.

Dalam beberapa tahun terakhir, metode pembelajaran mesin (machine learning) telah berkembang pesat sebagai solusi inovatif dalam bidang prediksi data lingkungan[5], [6]. Salah satu algoritma yang terbukti efektif untuk data deret waktu adalah Long Short-Term Memory (LSTM), yaitu bagian dari Recurrent Neural Network (RNN) yang dirancang untuk mempelajari pola jangka panjang[7], [8]. Berbagai studi sebelumnya telah menunjukkan keunggulan LSTM dibandingkan pendekatan tradisional seperti ARIMA dan regresi linier, terutama dalam menangkap dinamika temporal yang kompleks[9], [10], [11]. Namun demikian, sebagian besar penelitian tersebut masih menggunakan pendekatan univariat, sehingga kurang mampu menggambarkan interaksi antar faktor lingkungan secara menyeluruh[12].

Penelitian ini mengusulkan pemanfaatan model LSTM berbasis multivariat untuk meramalkan kualitas udara di Kota Surabaya. Pendekatan ini melibatkan penggunaan beberapa parameter lingkungan (suhu udara, kelembapan, dan kecepatan angin) secara simultan sebagai input untuk memprediksi dua target polutan utama, yaitu PM10 dan CO[4]. Dengan demikian, model yang dibangun diharapkan mampu memberikan prediksi yang lebih akurat dan representatif terhadap kondisi aktual[13], [14]. Tujuan utama dari penelitian ini adalah membangun sistem prediksi kualitas udara berbasis LSTM multivariat, menganalisis performanya, serta mengkaji efektivitasnya sebagai sistem pendukung pengambilan keputusan dalam pengelolaan kualitas udara perkotaan[15].

## II. KAJIAN TEORI

## A. Kualitas Udara

Kualitas udara merupakan indikator penting yang mencerminkan tingkat pencemaran di suatu wilayah. Parameter kualitas udara umumnya ditentukan oleh kadar polutan seperti PM10, PM2.5, CO, SO<sub>2</sub>, NO<sub>2</sub>, dan O<sub>3</sub>[16]. Dalam konteks penelitian ini, parameter PM10 (partikulat dengan diameter  $\leq 10$  mikron) dan CO (karbon monoksida) menjadi fokus utama karena keberadaannya yang dominan di wilayah perkotaan dan dampaknya yang signifikan terhadap kesehatan masyarakat. PM10 dapat masuk ke dalam saluran pernapasan manusia dan menyebabkan gangguan pernapasan, sedangkan CO adalah gas beracun yang mengganggu distribusi oksigen dalam darah. Pengukuran kualitas udara secara real-time dilakukan oleh Stasiun Pemantauan Kualitas Udara (SPKU) yang berada di berbagai titik, salah satunya di Kebonsari, Surabaya[17].

## B. Faktor-Faktor Meteorologi

Beberapa faktor meteorologi memiliki pengaruh langsung terhadap distribusi dan konsentrasi polutan udara[18]. Di antaranya adalah suhu udara, kelembapan udara, dan kecepatan angin[19].

- *Suhu udara* berpengaruh terhadap reaksi kimia atmosferik dan kerap dikaitkan dengan peningkatan emisi polutan tertentu.
- *Kelembapan* memengaruhi dispersi dan pelarutan polutan di udara.
- *Kecepatan angin* memengaruhi proses difusi dan transportasi polutan antar wilayah. Ketiga variabel ini digunakan dalam penelitian sebagai input multivariat karena perannya dalam membentuk pola distribusi PM10 dan CO secara temporal.

## C. Multivariat LSTM

Long Short-Term Memory (LSTM) merupakan salah satu bentuk jaringan saraf tiruan yang dirancang untuk mengatasi permasalahan dalam pemrosesan data deret waktu, terutama ketika terdapat ketergantungan jangka panjang. Kemampuannya dalam menyimpan informasi penting dalam jangka waktu tertentu menjadikan LSTM sebagai salah satu algoritma yang sangat efektif dalam berbagai aplikasi peramalan, termasuk untuk memodelkan data kualitas udara[9].

Deret waktu adalah rangkaian data yang dikumpulkan dan dicatat berdasarkan urutan waktu[16], [20]. Ciri khas data deret waktu adalah adanya ketergantungan antara nilai saat ini dengan nilai-nilai sebelumnya (autokorelasi)[21]. Dalam studi kualitas udara, deret waktu digunakan untuk memodelkan dan meramalkan konsentrasi polutan berdasarkan pola masa lalu[22]. Komponen penting dalam analisis deret waktu meliputi tren, musiman, dan fluktuasi acak[23].

## D. Long Short-Term Memory (LSTM)

LSTM merupakan pengembangan dari RNN yang mampu menangani ketergantungan jangka panjang dengan lebih baik[24], [25]. LSTM memiliki tiga gerbang utama yaitu *forget gate*, *input gate*, dan *output gate*, yang berfungsi mengatur aliran informasi yang relevan ke dalam dan ke luar memori[26]. Arsitektur ini membuat LSTM lebih stabil dan efisien dalam mempelajari pola data deret waktu yang kompleks seperti fluktuasi kualitas udara[27]. Dalam

penelitian ini[9], [10], LSTM digunakan sebagai model utama untuk meramalkan nilai PM10 dan CO berdasarkan

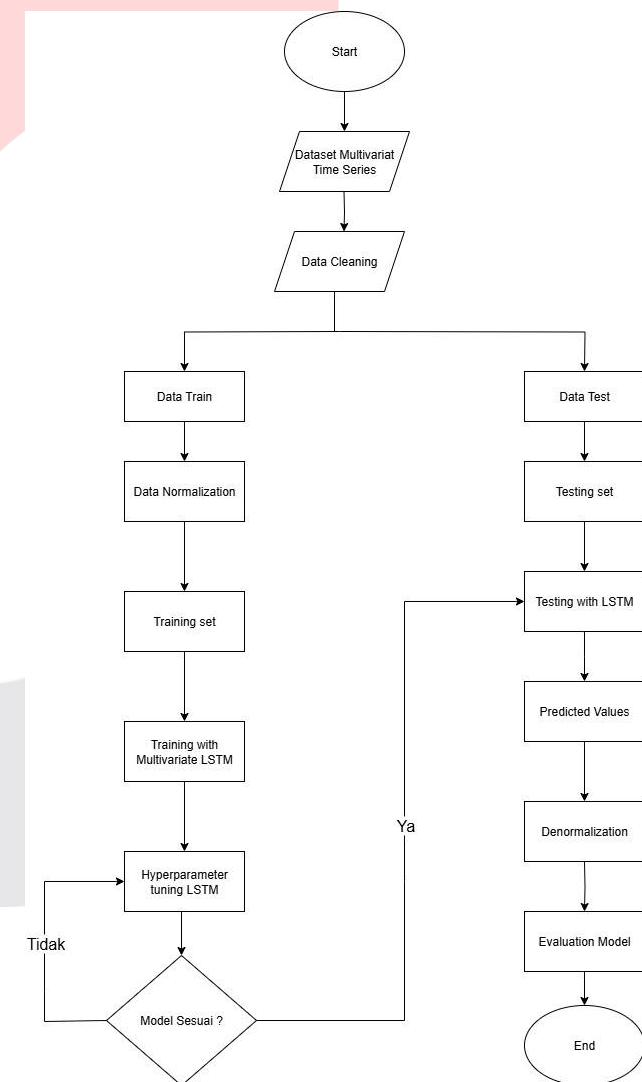
input multivariat.

## E. Evaluasi Kinerja Model

Evaluasi model dilakukan untuk mengukur seberapa akurat model prediktif dalam menebak nilai yang mendekati data aktual. Beberapa metrik evaluasi yang digunakan dalam penelitian ini adalah:

- *Root Mean Square Error (RMSE)*: Mengukur selisih rata-rata kuadrat antara nilai aktual dan prediksi.
- *Mean Absolute Percentage Error (MAPE)*: Mengukur kesalahan dalam bentuk persentase terhadap nilai aktual.
- *Mean Square Error (MSE)* dan *R-Square (R<sup>2</sup>)* juga digunakan untuk memperkuat analisis performa. Semakin kecil nilai RMSE dan MAPE, maka semakin baik performa model prediksi[28].

## III. METODE



GAMBAR 1  
(DIAGRAM ALIR MULTIVARIAT LSTM)

Penelitian ini pada gambar 1 mengenai floprosesnya dirancang untuk membangun dan mengevaluasi model peramalan kualitas udara berbasis Long Short-Term Memory (LSTM), yang merupakan pengembangan dari Recurrent Neural Network (RNN). Model yang digunakan bersifat multivariat, yaitu memanfaatkan beberapa variabel masukan

untuk memprediksi dua variabel target. Penelitian dilakukan dari Januari 2024 hingga Juli 2024 dengan pendekatan kuantitatif yang memanfaatkan data sekunder[2].

Data diperoleh dari Stasiun Pemantauan Kualitas Udara (SPKU) Kebonsari yang dikelola oleh Dinas Lingkungan Hidup (DLH) Kota Surabaya. SPKU merupakan sistem pengukuran kualitas udara secara otomatis yang mencatat data lingkungan secara berkala. Data yang digunakan meliputi lima variabel, yaitu:

- PM10 (Particulate Matter dengan diameter  $\leq 10$  mikron)
- CO (Carbon Monoxide)
- Suhu udara
- Kelembaban
- Kecepatan angin

Data dikumpulkan dengan frekuensi setiap satu jam dari

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Keterangan:

$y_i$  = nilai aktual

$\hat{y}_i$  = nilai prediksi

$n$  = jumlah data

Januari 2022 hingga Desember 2023, dan kemudian diolah untuk keperluan pemodelan. Berikut ini adalah tahapan-tahapan penelitian secara sistematis:

#### 1. Pengumpulan dan Praproses Data

Data yang diperoleh dari SPKU Kebonsari memiliki missing value dan ketidaksesuaian format waktu. Oleh karena itu, langkah awal adalah melakukan interpolasi linier untuk mengisi nilai yang hilang. Selanjutnya dilakukan normalisasi data ke dalam dua rentang: [0,1] dan [-1,1], sesuai karakteristik input dan target variabel. Normalisasi digunakan agar data numerik berada dalam skala yang sama sehingga mempercepat konvergensi model LSTM.

#### 2. Analisis Eksploratif dan Korelasi

Dilakukan analisis deskriptif dan visualisasi (boxplot dan heatmap) untuk mengidentifikasi pola musiman, outlier, dan korelasi antar variabel. Hasil analisis menunjukkan bahwa suhu, kelembaban, dan kecepatan angin memiliki hubungan yang cukup kuat terhadap PM10 dan CO.

#### 3. Pembangunan Model LSTM

LSTM adalah jenis jaringan saraf tiruan yang memiliki memori jangka panjang dan dapat menghindari masalah *vanishing gradient* pada RNN klasik. Arsitektur LSTM yang digunakan terdiri dari:

- Input layer (dengan 3 fitur: suhu, kelembaban, kecepatan angin)
- Dua hidden layer LSTM
- Dense output layer (untuk prediksi PM10 dan CO secara simultan)

Fungsi aktivasi tanh digunakan pada hidden layer, dan dropout layer ditambahkan untuk menghindari overfitting.

#### 4. Pembagian Data dan Pelatihan Model

Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%). Data dilatih dalam format sekuensial dengan time step = 24, artinya model mempelajari pola dari 24 jam sebelumnya untuk memprediksi nilai pada jam ke-25.

#### 5. Tuning Hyperparameter

Untuk mengoptimalkan kinerja model, dilakukan pencarian kombinasi terbaik terhadap jumlah unit neuron dan nilai dropout dengan menggunakan teknik Grid Search. Evaluasi dilakukan berdasarkan nilai metrik akurasi seperti RMSE dan MAPE.

#### 6. Evaluasi Kinerja Model

Kinerja model dievaluasi dengan beberapa metrik, antara lain:

- Root Mean Square Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Symmetric Mean Absolute Percentage Error (SMAPE)
- Mean Square Error (MSE)
- R-Squared ( $R^2$ )

Rumus untuk menghitung MAPE diberikan dalam Persamaan (1):

Semakin rendah nilai RMSE dan MAPE, maka semakin baik performa model.

## IV. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk membangun model peramalan kualitas udara berbasis Multivariate LSTM dengan mempertimbangkan variabel meteorologi utama sebagai input. Bagian ini menyajikan hasil eksplorasi data, transformasi dataset, pelatihan model, serta evaluasi performa peramalan terhadap parameter PM10 dan CO yang telah dilakukan secara sistematis.

#### A. Eksplorasi Data dan Karakteristik

- Penanganan Data Missing Value

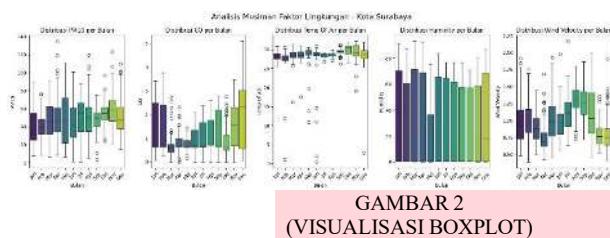
Tabel 1 terlihat bahwa data hilangnya berbeda-beda untuk setiap variabel. Variabel Kecepatan Angin memiliki jumlah data hilang terbanyak, yaitu 32, diikuti oleh PM10 (29 data), CO (20 data), Kelembaban (19 data), dan yang paling sedikit adalah Temperatur Udara (15 data). Perbedaan jumlah data hilang ini menunjukkan adanya tantangan dalam analisis data yang tidak merata di seluruh variabel. Untuk menagatasi data missing value menggunakan pendekatan interpolasi linier dianggap memadai karena data yang hilang tidak masif dan masih berada dalam interval harian yang berdekatan.

TABEL 1  
(JUMLAH DATA MISSING VALUE)

Variabel	Jumlah Missing Value
PM10	29
CO	20
Temp Of Air	15
Humidity	19
Wind Velocity	32

- Distribusi Musiman Parameter Kualitas Udara Berdasarkan Bulan

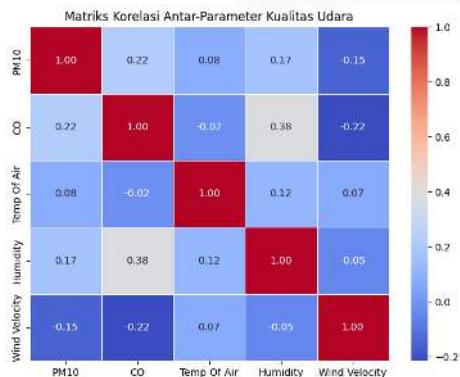
Visualisasi menunjukkan variasi maksimum beberapa parameter polusi dan cuaca di tiap stasiun. PM10 dan CO tercatat tinggi di beberapa lokasi, terutama yang padat kendaraan. Suhu relatif stabil antar wilayah. Jumlah kendaraan bervariasi signifikan dan berkorelasi dengan tingkat polusi. Kecepatan angin yang rendah di beberapa titik diduga menghambat dispersi polutan.



GAMBAR 2  
(VISUALISASI BOXPLOT)

### B. Analisis Korelasi

Berdasarkan matriks korelasi pada gambar 3, hubungan antar parameter kualitas udara umumnya lemah. CO menunjukkan korelasi sedang dengan kelembaban (0,38), dan rendah dengan PM10 (0,22). Kecepatan angin berkorelasi negatif terhadap CO dan PM10, mengindikasikan peran angin dalam menyebarluaskan polutan. Korelasi antar variabel lain seperti suhu dan kelembaban relatif kecil, sehingga masing-masing parameter memiliki pengaruh tersendiri terhadap kualitas udara.



GAMBAR 3  
(ANALISIS KORELASI)

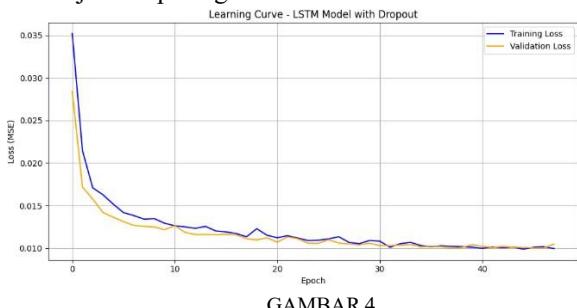
### C. Kinerja Model LSTM

Model LSTM dibangun untuk memprediksi PM10 dan CO berdasarkan data suhu, kelembapan, dan kecepatan angin selama 24 jam sebelumnya. Arsitektur terdiri dari input berformat [samples, 24, 3], satu lapisan LSTM dengan 64 unit, dan dense layer dengan dua neuron output. Model dikompilasi menggunakan MSE sebagai loss function, Adam sebagai optimizer, dan MAE sebagai metrik evaluasi.

- Visualisasi Akurasi dan Validasi

Evaluasi performa model dilakukan melalui grafik learning curve yang menunjukkan penurunan MSE pada data pelatihan dan validasi seiring bertambahnya epoch. Kurva keduanya menunjukkan tren menurun dan konvergen, tanpa selisih mencolok. Hal ini mengindikasikan bahwa

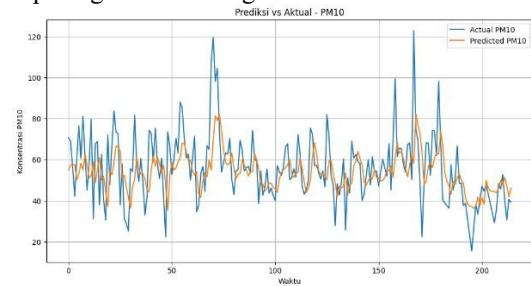
model tidak mengalami overfitting dan mampu mempelajari pola data secara seimbang. Penggunaan dropout terbukti efektif dalam menjaga kemampuan generalisasi model, visualisasi ditunjukkan pada gambar 4



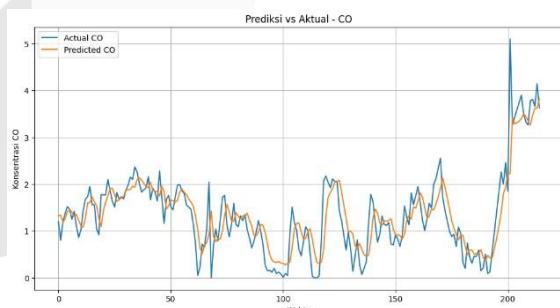
GAMBAR 4  
(TRAINING DAN VALIDATION MODEL LSTM)

- Visualisasi Prediksi dan Aktual

Prediksi PM10 menunjukkan kesesuaian dengan tren data aktual, meskipun cenderung melemahkan variasi ekstrem. Hal ini menunjukkan model mampu mengenali pola umum, namun kurang sensitif terhadap lonjakan tajam. Sementara itu, prediksi CO menunjukkan hasil yang lebih stabil dan akurat, dengan pola naik-turun yang mengikuti data aktual secara konsisten, meski masih terdapat sedikit deviasi pada titik-titik ekstrem. Hasil ditunjukkan pada gambar 5 dan gambar 6



GAMBAR 5  
(PREDIKSI DAN AKTUAL PM10 MODEL LSTM)



GAMBAR 6  
(PREDIKSI DAN AKTUAL CO MODEL LSTM)

- Hasil Evaluasi Model LSTM

Hasil Evaluasi Model LSTM pada table 2 menghasilkan MSE sebesar 111,77 dan RMSE 10,57, menunjukkan rata-rata kesalahan prediksi dalam skala asli. Nilai MAPE sebesar 54,55% menandakan deviasi yang cukup besar terhadap nilai aktual, sedangkan SMAPE 28,74% menunjukkan kesalahan relatif yang lebih seimbang. Hasil ini mengindikasikan bahwa model cukup mampu menangkap pola temporal, namun

masih kurang akurat pada data yang fluktuatif dan bernilai kecil, sehingga masih dapat ditingkatkan.

TABEL 2  
(HASIL EVALUASI KINERJA MODEL LSTM)

Variabel	Jumlah Missing Value
MSE	111.7745
RMSE	10.5723
MAPE	54.55%
SMAPE	28.74%

#### D. Tuning Hyperparameter

Pemilihan hyperparameter menjadi kunci dalam membangun model LSTM yang optimal. Proses tuning dilakukan dengan menguji berbagai kombinasi batch size, dropout, epoch, L2 regularisasi, dan jumlah unit LSTM. Evaluasi dilakukan berdasarkan nilai loss pada data validasi. Tabel 3 memuat hasil eksplorasi, sementara Tabel 4 menunjukkan konfigurasi terbaik yang kemudian digunakan untuk pelatihan akhir karena memberikan performa prediktif tertinggi.

TABEL 3  
(HASIL EVALUASI TUNING HYPERPARAMETER LSTM)

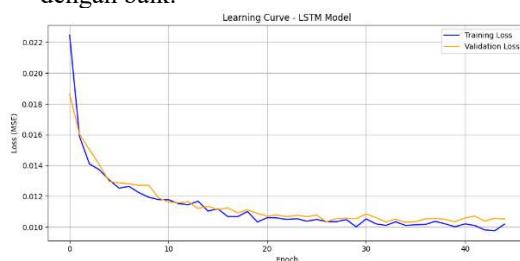
Model	Batch Size	Dropout	Epochs	Learning Rate	Unit	Val Loss
1.	16	0.2	50	0.0001	32	0.0105
2.	16	0.2	50	0.0001	64	0.0104
3	16	0.2	50	0.001	128	0.0104
...	...	...	...	...	...	...
36	32	0.5	50	0.001	128	0.0118

TABEL 4  
(BEST TUNING HYPERPARAMETER LSTM)

Variabel	Jumlah Missing Value
Batch Size	16
Dropout Rate	0.3
Epochs	50
Learning Rate	0.0001
Jumlah Unit LSTM	128

- Visualisasi Akurasi dan Validasi Tuning

Grafik pada gambar 7, menunjukkan Kurva pelatihan menunjukkan penurunan error yang stabil tanpa indikasi overfitting, menandakan model belajar efektif dan mampu menggeneralisasi data dengan baik.



GAMBAR 7

(TRAINING DAN VALIDATION TUNING HYPERPARAMETER)

- Hasil Evaluasi Tuning Hyperparameter

Setelah tuning hyperparameter, model LSTM menghasilkan MSE sebesar 113,62 dan RMSE 10,66. Nilai MAPE tercatat 49,45%, sementara SMAPE 28,17%, menunjukkan

penurunan kesalahan persentase dibanding sebelum tuning. Meski peningkatan pada MSE dan RMSE relatif kecil, hasil ini mencerminkan peningkatan stabilitas dan efisiensi model dalam memprediksi data.

TABEL 5  
(HASIL EVALUASI KINERJA TUNING MODEL LSTM)

Variabel	Jumlah Missing Value
MSE	111.7745
RMSE	10.5723
MAPE	54.55%
SMAPE	28.74%

#### V. KESIMPULAN

Penelitian ini menunjukkan bahwa variabel meteorologi seperti suhu, kelembapan, dan kecepatan angin memiliki pengaruh terhadap kualitas udara, khususnya pada konsentrasi PM10 dan CO. Analisis korelasi mengonfirmasi adanya hubungan signifikan, menandakan pentingnya faktor cuaca dalam prediksi polusi udara. Model LSTM multivariat berhasil dikembangkan untuk memprediksi kualitas udara di Surabaya, dengan tahapan mencakup normalisasi, pembentukan data sekuensial, dan tuning hyperparameter. Hasil evaluasi menunjukkan performa yang cukup baik dengan MSE 113,62, RMSE 10,66, MAPE 49,45%, dan SMAPE 28,17%. Nilai-nilai ini menandakan akurasi prediksi yang memadai dan kemampuan model dalam mengenali pola temporal. Dengan demikian, LSTM memiliki potensi untuk diterapkan sebagai alat bantu prediktif dalam sistem pemantauan kualitas udara.

#### REFERENSI

- [1] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, “Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor,” *JIPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 7, no. 2, pp. 496–507, May 2022, doi: 10.29100/jipi.v7i2.2843.
- [2] T. V. Damayanti and R. E. Handriyono, “Monitoring Kualitas Udara Ambien Melalui Stasiun Pemantau Kualitas Udara Wonorejo, Kebonsari Dan Tandes Kota Surabaya,” *Environ. Eng. J. ITATS*, vol. 2, no. 1, pp. 11–18, Mar. 2022, doi: 10.31284/j.envitats.2022.v2i1.2897.
- [3] M. N. Atalla and F. Agustina, “Penentuan Daerah Rawan Polusi Udara Menggunakan Logika Fuzzy Metode Mamdani”.
- [4] J. A. Ramírez-Montañez, J. D. J. Rangel-Magdaleno, M. A. Aceves-Fernández, and J. M. Ramos-Arreguín, “Modeling of Particulate Pollutants Using a Memory-Based Recurrent Neural Network Implemented on an FPGA,” *Micromachines*, vol. 14, no. 9, p. 1804, Sep. 2023, doi: 10.3390/mi14091804.
- [5] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, “A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks,” *Future Internet*, vol. 15, no. 8, p. 255, Jul. 2023, doi: 10.3390/fi15080255.
- [6] G. Cican, A.-N. Buturache, and R. Mirea, “Applying Machine Learning Techniques in Air Quality Prediction—A Bucharest City Case Study,” *Sustainability*,

vol. 15, no. 11, p. 8445, May 2023, doi: 10.3390/su15118445.

[7] Y. Chen, S. Cui, P. Chen, Q. Yuan, P. Kang, and L. Zhu, "An LSTM-based neural network method of particulate pollution forecast in China," *Environ. Res. Lett.*, vol. 16, no. 4, p. 044006, Apr. 2021, doi: 10.1088/1748-9326/abe1f5.

[8] A. Azzouni and G. Pujolle, "A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction," Jun. 08, 2017, *arXiv: arXiv:1705.05690*. doi: 10.48550/arXiv.1705.05690.

[9] M. A. A. Bakar, N. M. Ariff, M. S. Mohd Nadzir, O. L. Wen, and F. N. A. Suris, "Prediction of Multivariate Air Quality Time Series Data using Long Short-Term Memory Network," *Malays. J. Fundam. Appl. Sci.*, vol. 18, no. 1, pp. 52–59, Feb. 2022, doi: 10.11113/mjfas.v18n1.2393.

[10] Muh. A. Faishol, E. Endroyono, and A. N. Irfansyah, "PREDICT URBAN AIR POLLUTION IN SURABAYA USING RECURRENT NEURAL NETWORK – LONG SHORT TERM MEMORY," *JUTI J. Ilm. Teknol. Inf.*, pp. 102–114, Jul. 2020, doi: 10.12962/j24068535.v18i2.a988.

[11] R. T. D. Kurniawati, R. Rahmawati, and Y. Wilandari, "PENGELO MPOKAN KUALITAS UDARA AMBIEN MENURUT KABUPATEN/KOTA DI JAWA TENGAH MENGGUNAKAN ANALISIS KLASTER".

[12] F. M. Ottaviani and A. D. Marco, "Multiple Linear Regression Model for Improved Project Cost Forecasting," *Procedia Comput. Sci.*, vol. 196, pp. 808–815, 2022, doi: 10.1016/j.procs.2021.12.079.

[13] R. T. Alegado and G. M. Tumbay, "Statistical and Machine Learning Methods for Vaccine Demand Forecasting: A Comparative Analysis," *J. Comput. Commun.*, vol. 08, no. 10, pp. 37–49, 2020, doi: 10.4236/jcc.2020.810005.

[14] V. Papastefanopoulos, P. Linardatos, T. Panagiotakopoulos, and S. Kotsiantis, "Multivariate Time-Series Forecasting: A Review of Deep Learning Methods in Internet of Things Applications to Smart Cities," *Smart Cities*, vol. 6, no. 5, pp. 2519–2552, Sep. 2023, doi: 10.3390/smartcities6050114.

[15] M. E. S. Wicaksono, G. M. Arya Sasmita, and I. P. A. Eka Pratama, "Peramalan Kualitas Udara di Kota Jakarta Pusat dengan Metode Long Short-Term Memory dan Support-Vector Regression," *JITTER J. Ilm. Teknol. Dan Komput.*, vol. 4, no. 1, p. 1576, Jan. 2023, doi: 10.24843/JTRTI.2023.v04.i01.p03.

[16] S. D. Pasupuleti and S. Ludwig, "Forecasting Air Quality: A Comparative Study of Time Series Approaches," Sep. 07, 2023, *Preprints*. doi: 10.22541/au.169406670.05499290/v1.

[17] K. Fahim, S. A. Sholichah, L. Hanafi, S. D. Surjanto, I. Herisman, and W. F. Doctorina, "Prediksi Indeks Standar Pencemar Udara (ISPU) di Kota Surabaya Menggunakan Rantai Markov Waktu Diskrit," *J. Ilm. Soulmath J. Edukasi Pendidik. Mat.*, vol. 12, no. 2, pp. 121–134, Aug. 2024, doi: 10.25139/smj.v12i2.8684.

[18] G. Naresh and D. B. Indira, "Air Pollution Prediction using Multivariate LSTM Deep Learning Model".

[19] A. P. Baluk, H. Yasin, and Sugito, "Peramalan Tinggi Gelombang Laut Dengan Metode Vector Autoregressive-Radial Basis Function Network (Var-Rbfm)," *J Stat. J. Ilm. Teori Dan Apl. Stat.*, vol. 13, no. 2, pp. 39–46, Dec. 2020, doi: 10.36456/jstat.vol13.no2.a3270.

[20] H. Hajmohammadi and B. Heydecker, "Multivariate time series modelling for urban air quality," *Urban Clim.*, vol. 37, p. 100834, May 2021, doi: 10.1016/j.ulim.2021.100834.

[21] R. Espinosa, J. Palma, F. Jiménez, J. Kamińska, G. Sciaivocco, 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.

[22] B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, "Forecasting air quality time series using deep learning," *J. Air Waste Manag. Assoc.*, vol. 68, no. 8, pp. 866–886, Aug. 2018, doi: 10.1080/10962247.2018.1459956.

[23] T. Mathonsi and T. L. Van Zyl, "A Statistics and Deep Learning Hybrid Method for Multivariate Time Series Forecasting and Mortality Modeling," *Forecasting*, vol. 4, no. 1, pp. 1–25, Dec. 2021, doi: 10.3390/forecast4010001.

[24] R. Navares and J. L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models," *Ecol. Inform.*, vol. 55, p. 101019, Jan. 2020, doi: 10.1016/j.ecoinf.2019.101019.

[25] Stavelin Abhinandithe K, Madhu B, Somanathan Balasubramanian, and Sridhar Ramachandran, "Forecasting Multivariate time-series data using LSTM Neural Network in Mysore district, Karnataka," *Indian J. Public Health Res. Dev.*, vol. 13, no. 4, pp. 284–289, Oct. 2022, doi: 10.37506/ijphrd.v14i4.18631.

[26] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting," *Procedia Comput. Sci.*, vol. 135, pp. 89–98, 2018, doi: 10.1016/j.procs.2018.08.153.

[27] Y. Yin and P. Shang, "Forecasting traffic time series with multivariate predicting method," *Appl. Math. Comput.*, vol. 291, pp. 266–278, Dec. 2016, doi: 10.1016/j.amc.2016.07.017.

[28] A. W. Saputra, A. P. Wibawa, U. Pujiyanto, A. B. Putra Utama, and A. Nafalski, "LSTM-based Multivariate Time-Series Analysis: A Case of Journal Visitors Forecasting," *Ilk. J. Ilm.*, vol. 14, no. 1, pp. 57–62, Apr. 2022, doi: 10.33096/ilkom.v14i1.1106.57-62.

•