Jurnal ilmiah Sistem Informasi dan Ilmu Komputer
Volume. 6 Nomor. 1 Maret 2026
e-ISSN: 2827-7953; p-ISSN: 2827-8135, Hal. 616-629
DOI: https://doi.org/10.55606/juisik.v6i1.2228
Tersedia: https://journal.sinov.id/index.php/juisik

Analisis Model Long Short-Term Memory (LSTM) untuk Prediksi Kadar Particulate Matter 2.5

Vella Puspitasari Wijayanti[1*], Adie Wahyudi Oktavia Gama[2]
[1-2] Teknologi Informasi, Universitas Pendidikan Nasional, Indonesia
Email: vellapuspita0@gmail.com [1], adiewahyudi@undiknas.ac.id [2]*

**Penulis Korespondensi: vellapuspita0@gmail.com [1]*

Abstract. *Particulate Matter 2.5 (PM2.5) is a hazardous air pollutant that can adversely affect human health and the environment. Accurate prediction of PM2.5 concentrations is essential to support air quality monitoring and pollution mitigation policy-making. This study aims to analyze the performance of the Long Short-Term Memory (LSTM) model in predicting PM2.5 concentrations based on time series data. The LSTM model, as a variant of Recurrent Neural Networks (RNN), is selected due to its ability to capture long-term dependencies in sequential data and to overcome the vanishing gradient problem commonly encountered in conventional RNN models. This study utilizes historical PM2.5 concentration data along with supporting variables such as temperature, humidity, wind speed, and other pollutant parameters. The research process includes data preprocessing, determination of the LSTM architecture (number of layers, hidden units, timestep), model training, and evaluation using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The results indicate that the LSTM model is capable of providing accurate predictions for short- to medium-term forecasting horizons, with performance superior to baseline models such as ARIMA and simple RNN. This study is expected to serve as a reference for the development of more reliable and practical deep learning-based air quality prediction systems in urban areas.*

Keywords: *Air Quality; Deep Learning; PM2.5; Recurrent Neural Networks; Time Series Prediction.*

Abstrak. Particulate Matter 2.5 (PM2.5) merupakan polutan udara berbahaya yang dapat memengaruhi kesehatan manusia dan lingkungan. Prediksi kadar PM2.5 secara akurat sangat penting untuk mendukung pemantauan kualitas udara dan pengambilan kebijakan mitigasi pencemaran. Penelitian ini bertujuan untuk menganalisis performa model Long Short-Term Memory (LSTM) dalam memprediksi kadar PM2.5 berbasis data time series. Model LSTM, sebagai varian Recurrent Neural Network (RNN), dipilih karena kemampuannya menangkap ketergantungan jangka panjang pada data sekuensial, mengatasi permasalahan vanishing gradient yang sering terjadi pada model RNN konvensional. Penelitian ini menggunakan data historis kadar PM2.5 beserta variabel pendukung seperti suhu, kelembaban, kecepatan angin, dan parameter polutan lainnya. Proses penelitian meliputi pra-pemrosesan data, penentuan arsitektur LSTM (jumlah layer, unit hidden, timestep), pelatihan model, serta evaluasi menggunakan metrik Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2). Hasil analisis menunjukkan bahwa model LSTM mampu memberikan prediksi yang akurat untuk horizon waktu pendek hingga menengah, dengan performa yang lebih baik dibandingkan model baseline seperti ARIMA atau RNN sederhana. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem prediksi kualitas udara berbasis deep learning yang lebih andal dan aplikatif di wilayah perkotaan.

Kata Kunci: Jaringan Saraf Berulang; Kualitas Udara; Pembelajaran Mendalam; PM2.5; Prediksi Deret Waktu.

1. PENDAHULUAN

Pencemaran udara kini menjadi salah satu ancaman kesehatan masyarakat paling serius di tingkat global. Di antara berbagai polutan, Particulate Matter 2.5 ($PM_{2.5}$) mendapat perhatian khusus karena ukurannya yang sangat kecil (< 2,5 mikrometer). Ukuran ini memungkinkan partikel tersebut menembus pertahanan sistem pernapasan manusia dan masuk ke dalam sistem peredaran darah secara langsung, yang dalam jangka panjang meningkatkan risiko penyakit kardiovaskular, stroke, hingga kanker paru-paru. Partikel ini mudah menembus sistem

Naskah Masuk: 15 Januari 2026; Revisi: 19 Februari 2026; Diterima: 16 Maret 2026; Terbit: 31 Maret 2026

pernapasan dan langsung masuk ke aliran darah, sehingga meningkatkan risiko penyakit kardiovaskular, stroke, serta kanker paru-paru dalam jangka panjang. Kemampuan memprediksi konsentrasi $PM_{2.5}$ secara akurat menjadi sangat penting untuk mendukung kebijakan kesehatan publik dan sistem peringatan dini di kota-kota besar (Kristiani et al., 2022).

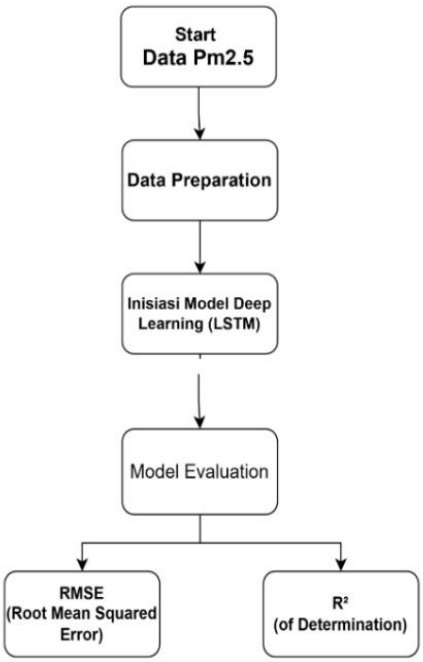
Dampak paparan $PM_{2.5}$ terhadap kesehatan telah dibuktikan melalui banyak penelitian epidemiologi internasional. Paparan jangka panjang dapat memicu penyakit pernapasan kronis, kanker paru-paru, dan gangguan kardiovaskular. Di Indonesia, peningkatan kasus ISPA dan asma di wilayah perkotaan kerap beriringan dengan lonjakan $PM_{2.5}$ pada musim kemarau. Selain itu, polutan ini juga merusak ekosistem, antara lain menurunkan kesuburan tanah dan mengganggu pertumbuhan vegetasi. Kelompok rentan seperti anak-anak, lansia, dan penderita penyakit kronis paling terdampak. Prediksi dini diharapkan dapat mengurangi risiko melalui peringatan awal dan pembatasan aktivitas luar ruangan (Bai et al., 2024).

Pendekatan statistik klasik seperti ARIMA Hasibuan, M. (2025) dan regresi linier sering kali kurang mampu menangani karakteristik data $PM_{2.5}$ yang bersifat non-linear dan memiliki ketergantungan temporal jangka panjang (Nourmohammad & Rashidi, 2025). Data ini sangat fluktuatif karena dipengaruhi faktor meteorologi seperti suhu, kelembaban, dan kecepatan angin. Keterbatasan tersebut mendorong penggunaan metode machine learning dan deep learning. Meskipun Recurrent Neural Network (RNN) dapat memproses data sekuensial, model ini rentan mengalami vanishing gradient pada data panjang (Nourmohammad & Rashidi, 2025), *(*Wang *et al., 2023)*. Oleh karena itu, Long Short-Term Memory (LSTM) Hadi, N., Alam, S., & Kurniawan, I. (2025) menjadi pilihan yang lebih unggul karena mampu menyimpan informasi jangka panjang melalui mekanisme gerbang (gate) (Ho *et al*.,2023

Long Short-Term Memory (LSTM) Wahyutama, S., Rismawan, T., & Bahri, S. (2026) merupakan arsitektur deep learning yang dirancang khusus untuk memproses data sekuensial dan time series. Model ini mengatasi kelemahan RNN dengan menambahkan tiga gerbang utama, yaitu forget gate, input gate, dan output gate. Kemampuan LSTM dalam menangkap pola temporal jangka panjang telah terbukti efektif dalam berbagai aplikasi prediksi lingkungan. Dalam konteks prediksi PM2.5 Guo, Z., Wu, S., Zhu, M., & He, G. (2025), LSTM dapat memanfaatkan data historis polutan dan variabel meteorologi secara simultan. Beberapa penelitian sebelumnya telah menunjukkan superioritas LSTM dibandingkan model baseline lainnya. Penerapan LSTM diharapkan dapat meningkatkan akurasi prediksi dan mendukung sistem pemantauan kualitas udara yang lebih andal di Indonesia (S. Basith *et al.*, 2022).

Beberapa penelitian terkini telah mengaplikasikan LSTM untuk prediksi PM2.5 dengan hasil yang menjanjikan. Kristiani et al. (2022) berhasil memprediksi PM2.5 jangka pendek menggunakan LSTM dengan RMSE yang rendah pada data *time series*. Sitompul (2024) menerapkan LSTM di wilayah Kalimantan Barat dan mendapatkan akurasi prediksi hingga 92% untuk kategori kualitas udara baik. Ho et al. (2022) membandingkan LSTM dengan model CMAQ di Korea dan menemukan bahwa LSTM memberikan skill forecast yang setara dengan prediksi manusia. Bai et al. (2024) mengintegrasikan CNN-LSTM dan mencapai R^2 sebesar 0,91 pada data Qingdao. Penelitian ini melanjutkan tren tersebut dengan menganalisis model LSTM secara mendalam untuk konteks Indonesia.

2. METODE



Gambar 1. Diagram Alur Penelitian.

Penelitian ini mengikuti alur yang sistematis dan terstruktur sebagaimana ditunjukkan pada Gambar 1 Proses dimulai dari Start Data PM2.5, yaitu pengumpulan data historis kadar Particulate Matter 2.5 beserta variabel pendukung seperti parameter meteorologi. Selanjutnya dilakukan tahap Data Preparation yang mencakup pembersihan data (data cleaning), penanganan missing value, normalisasi, serta pembagian data menjadi data latih (training) dan data uji (testing). Setelah data siap, penelitian melanjutkan ke tahap Inisiasi Model Deep Learning (LSTM), di mana arsitektur Long Short-Term Memory dirancang dengan

menentukan jumlah layer, unit hidden, timestep, fungsi aktivasi, optimizer, serta hyperparameter lainnya sebelum proses pelatihan model dilakukan.

Penelitian Pada tahap akhir, dilakukan Model Evaluation untuk menilai kinerja model LSTM yang telah dilatih. Evaluasi dilakukan menggunakan dua metrik utama, yaitu RMSE (Root Mean Squared Error) yang mengukur tingkat kesalahan prediksi secara rata-rata dan R^2 (Coefficient of Determination) yang menunjukkan seberapa baik model mampu menjelaskan variasi data aktual. Hasil evaluasi dari kedua metrik ini menjadi dasar untuk menentukan akurasi dan keandalan model dalam memprediksi kadar PM2.5. Diagram alur ini menggambarkan bahwa penelitian bersifat linier dengan evaluasi sebagai titik akhir yang menentukan kualitas hasil prediksi secara keseluruhan.

Tahapan Pra-pemrosesan Data Teknis

Tahap pra-pemrosesan merupakan langkah krusial dalam siklus pengembangan model *Deep Learning*, terutama pada algoritma *Deep Learning* seperti LSTM yang sangat sensitif terhadap kualitas dan skala data input. Proses ini bertujuan untuk mentransformasi data mentah yang bersifat tidak teratur menjadi format yang optimal bagi model agar dapat mengekstraksi fitur temporal secara efektif. Adapun tahapan detailnya adalah sebagai berikut:

Pembersihan Data (*Data Cleaning*): Menggunakan teknik interpolasi linear untuk mengisi nilai yang hilang (*missing values*) akibat kegagalan sensor, guna menjaga kontinuitas tren alami polutan $PM_{2.5}$ tanpa menghilangkan informasi penting.

Prediksi dengan Algoritma LSTM : Penelitian ini menggunakan algoritma Long Short-Term Memory (LSTM) karena efektivitasnya dalam memproses data *time-series* dengan ketergantungan jangka panjang, seperti fluktuasi $PM_{2.5}$. Melalui mekanisme gating, LSTM mampu mengelola informasi pada *memory cell* secara selektif untuk menghasilkan prediksi yang akurat.

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{n} L(y_i, \widehat{y_i}) + \lambda R(\theta)$$

Keterangan: (a) $j(\theta)$: *Total Cost Function..*, (b) L: Loss function (misal MSE), (c) $\lambda R(\theta)$: Penalti regularisasi $L2$..

Pembentukan *Sliding Window*: Mentransformasi data deret waktu menjadi struktur *supervised learning* dengan *time step* 30. Model menggunakan data historis 30 jam sebelumnya sebagai fitur input X untuk memprediksi konsentrasi $PM_{2.5}$ pada 1 jam berikutnya sebagai target y.

Arsitektur Model LSTM

Model *Deep Learning* yang dikembangkan dalam penelitian ini menggunakan arsitektur *Recurrent Neural Network* (RNN) varian *Long Short-Term Memory* (LSTM). Pemilihan arsitektur ini didasarkan pada kemampuannya yang superior dalam menangkap ketergantungan sekuensial dan pola jangka panjang pada data deret waktu polusi udara tanpa mengalami masalah *vanishing gradient*. Struktur detail dari model yang dirancang adalah sebagai berikut:

Lapisan Input: Lapisan ini berfungsi untuk menerima matriks data yang telah ditransformasi melalui proses *sliding window*. Data input memiliki dimensi yang sesuai dengan jumlah *time step* (30) dan jumlah fitur (1), yang merepresentasikan urutan konsentrasi $PM_{2.5}$ pada periode waktu sebelumnya.

Lapisan Tersembunyi (*Hidden Layers*): Arsitektur ini menggunakan strategi *Stacked LSTM* yang terdiri dari dua lapisan LSTM berturut-turut. Setiap lapisan memiliki 64 unit tersembunyi (*hidden units*). Penggunaan lapisan ganda ini bertujuan untuk meningkatkan kapasitas model dalam mempelajari representasi fitur yang lebih abstrak dan menangkap dependensi temporal yang lebih dalam dibandingkan hanya menggunakan satu lapisan tunggal.

Mekanisme Regulasi (*Dropout*): Untuk memitigasi risiko *overfitting*, di mana model terlalu menghafal data latih namun gagal melakukan generalisasi pada data uji, diterapkan lapisan *Dropout* dengan laju 0.2 di antara lapisan LSTM. Mekanisme ini bekerja dengan cara menonaktifkan unit neuron secara acak sebesar 20% selama proses pelatihan, sehingga memaksa model untuk mempelajari fitur yang lebih tangguh dan tidak bergantung pada neuron tertentu saja.

Lapisan Output: Pada tahap akhir, keluaran dari lapisan LSTM diteruskan ke sebuah lapisan *Dense* (lapisan terkoneksi penuh) dengan satu unit neuron. Lapisan ini menggunakan fungsi aktivasi linear untuk menghasilkan prediksi nilai tunggal yang bersifat kontinu, yaitu estimasi konsentrasi kadar $PM_{2.5}$ dalam satuan $\mu g/m^3$ untuk satu jam ke depan.

Metrik Evaluasi Performa

Untuk mengukur akurasi dan keandalan model *hybrid* LSTM dalam memprediksi kadar $PM_{2.5}$, digunakan dua metrik evaluasi regresi utama:

Root Mean Squared Error (RMSE): Metrik ini menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan prediksi. RMSE memberikan gambaran besarnya kesalahan dalam satuan yang sama dengan data asli $\mu g/m^3$ nilai yang lebih kecil menunjukkan akurasi model yang lebih presisi.

Coefficient of Determination (R^2): Metrik ini digunakan untuk menilai sejauh mana model mampu menjelaskan variabilitas data target. Nilai R^2 berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang sangat kuat dan akurat terhadap fluktuasi data.

3. HASIL DAN PEMBAHASAN

Dataset dan Pra-pemrosesan Data

Dataset yang digunakan dalam penelitian ini berasal dari stasiun pemantauan Denpasar Lumintang yang mencatat konsentrasi $PM_{2.5}$. Sebelum diumpankan ke model, data melalui proses normalisasi menggunakan *Min-Max Scaling* untuk memastikan semua fitur berada dalam rentang [0, 1]. Transformasi *Sliding Window* dengan *time step* sebesar 30 jam diterapkan untuk membentuk fitur input X dan target y, sehingga model dapat menangkap dependensi temporal secara efektif.

```
Data Awal:
     Tanggal     Waktu  Denpasar  Badung  Gianyar  Tabanan  Bangli  Klungkung  \
0 2025-01-01  00:00:00      21.8    38.0     38.0     38.0    38.0       21.2
1 2025-01-01  01:00:00      21.8    39.5     39.5     39.5    39.5       20.5
2 2025-01-01  02:00:00      20.0    38.8     38.8     38.8    38.8       20.3
3 2025-01-01  03:00:00      20.6    41.0     41.0     41.0    41.0       20.3
4 2025-01-01  04:00:00      23.2    44.4     44.4     44.4    44.4       20.2

   Karangasem  Singaraja  Jembrana
0        21.2       29.5      21.8
1        20.5       29.9      21.8
2        20.3       29.7      20.0
3        20.3       29.6      20.6
4        20.2       29.9      23.2
```

Gambar 2. *Preprocessing.*

Proses pra-pemrosesan data merupakan tahapan krusial yang dilakukan untuk mentransformasikan data mentah kualitas udara menjadi format yang optimal bagi model Deep Learning. Tahapan ini diawali dengan pembersihan data melalui teknik interpolasi linear untuk menangani nilai yang hilang (*missing values*) guna menjaga kontinuitas deret waktu tanpa kehilangan informasi penting. Selanjutnya, dilakukan normalisasi menggunakan metode *Min-Max Scaling* untuk menyamakan skala seluruh fitur ke dalam rentang 0 hingga 1, sehingga mempercepat proses konvergensi saat pelatihan model LSTM. Terakhir, data diorganisasikan menggunakan metode *Sliding Window* dengan *time step* sebesar 30, yang berfungsi mengubah urutan data deret waktu menjadi struktur *supervised learning*. Melalui rangkaian prosedur ini, data yang dihasilkan memiliki korelasi temporal yang kuat, sehingga memudahkan model dalam menangkap pola fluktuasi konsentrasi $PM_{2.5}$ di masa depan berdasarkan data historis sebelumnya.

Hasil Evaluasi Model Baseline (LSTM Standar)

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 30, 64)	16,896
dropout_1 (Dropout)	(None, 30, 64)	0
lstm_3 (LSTM)	(None, 64)	33,024
dense_1 (Dense)	(None, 1)	65

Total params: 49,985 (195.25 KB)
Trainable params: 49,985 (195.25 KB)
Non-trainable params: 0 (0.00 B)

Gambar 3. Model Baseline LSTM.

Tahap pembangunan model dilakukan dengan merancang arsitektur *Recurrent Neural Network* jenis *Long Short-Term Memory* (LSTM) yang dikonfigurasi secara berlapis untuk menangkap dependensi temporal pada data $PM_{2.5}$. Arsitektur ini terdiri dari lapisan input yang menerima matriks data sesuai dengan *time step* yang telah ditentukan, diikuti oleh dua lapisan tersembunyi (*hidden layers*) yang masing-masing memiliki 64 unit LSTM untuk mengekstraksi fitur sekuensial secara mendalam. Untuk menjaga generalisasi model dan meminimalisir risiko *overfitting*, diterapkan mekanisme regulasi berupa lapisan *Dropout* dengan laju 0.2 di antara lapisan LSTM, yang berfungsi memutus koneksi neuron secara acak selama proses pelatihan. Rangkaian arsitektur ini diakhiri dengan sebuah lapisan *Dense* sebagai lapisan output yang menggunakan aktivasi linear untuk menghasilkan nilai prediksi kontinu dari konsentrasi polutan udara.

```
[24]
✓ 1d
    # 8. Prediksi dan Evaluasi

    y_pred = model.predict(X_test)
    y_pred_rescaled = scaler.inverse_transform(y_pred)
    y_test_rescaled = scaler.inverse_transform(y_test.reshape(-1, 1))

    rmse = math.sqrt(mean_squared_error(y_test_rescaled, y_pred_rescaled))
    r2 = r2_score(y_test_rescaled, y_pred_rescaled)

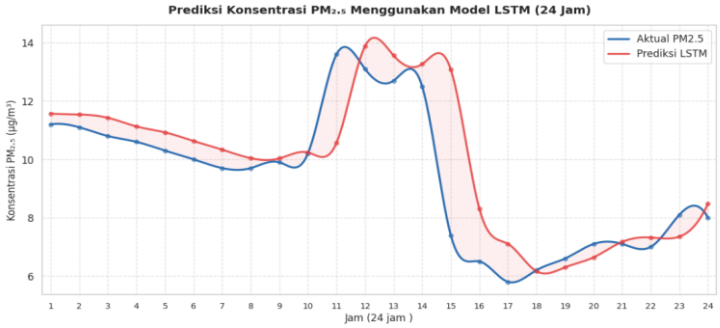
    print(f"RMSE: {rmse:.2f}")
    print(f"R²: {r2:.3f}")

    22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 23ms/step
    RMSE: 2.27
    R²: 0.729
```

Gambar 4. Penerapan Prediksi dan Evaluasi LSTM.

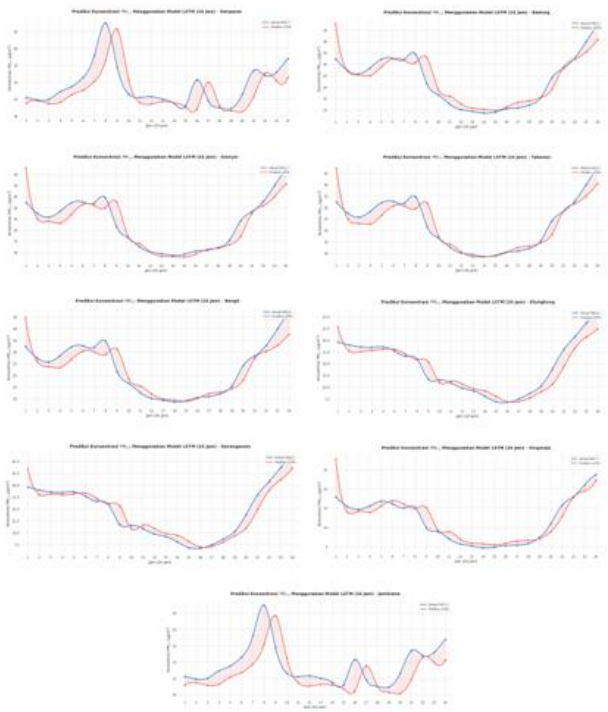
Tahap prediksi dan evaluasi merupakan proses final untuk menguji performa model LSTM yang telah dilatih menggunakan data uji (*testing data*). Proses ini diawali dengan melakukan pembalikan normalisasi (*inverse transform*) pada hasil prediksi model guna mengembalikan nilai skala 0–1 ke dalam satuan asli konsentrasi $PM_{2.5}$ yaitu $\mu g/m^3$, sehingga hasil dapat diinterpretasikan secara fisik. Evaluasi dilakukan secara kuantitatif menggunakan dua metrik utama, yaitu *Root Mean Squared Error* (RMSE) untuk mengukur rata-rata besarnya penyimpangan antara nilai prediksi dan nilai aktual, serta *Coefficient of Determination* (R^2) untuk mengukur sejauh mana model mampu menjelaskan variabilitas data. Selain evaluasi

numerik, dilakukan juga visualisasi kurva perbandingan antara data aktual dan hasil prediksi untuk melihat kemampuan model dalam mengikuti pola fluktuasi polutan udara dari waktu ke waktu. Melalui integrasi metrik ini, tingkat akurasi dan keandalan model dalam memprediksi kualitas udara dapat divalidasi secara objektif.



Gambar 5. Hasil Model Prediksi $PM_{2.5}$.

Grafik hasil prediksi ini merupakan representasi visual yang membandingkan performa model LSTM terhadap data aktual pada tahap pengujian. Kurva yang berwarna biru melambangkan nilai aktual konsentrasi $PM_{2.5}$ yang diamati di lapangan, sedangkan kurva berwarna merah menunjukkan nilai yang diprediksi oleh model. Kedekatan pola antara kedua kurva tersebut mengindikasikan kemampuan model dalam menangkap fluktuasi dan tren data deret waktu secara akurat. Selain itu, adanya area bayangan atau selisih tipis di antara kedua garis tersebut merepresentasikan *error* atau residu dari prediksi, yang secara kuantitatif diukur melalui nilai RMSE. Hasil visualisasi ini mempertegas bahwa model tidak hanya mampu memprediksi nilai rata-rata, tetapi juga cukup responsif terhadap perubahan konsentrasi polutan yang bersifat dinamis, sehingga valid untuk digunakan sebagai alat pemantauan kualitas udara.



Gambar 6. Kurva Perbandingan PM2.5 Antar Wilayah.

Tahap pembangunan model dilakukan dengan merancang arsitektur *Recurrent Neural Network* jenis *Long Short-Term Memory* (LSTM) yang dikonfigurasi secara berlapis untuk menangkap dependensi temporal pada data $PM_{2.5}$. Arsitektur ini terdiri dari dua lapisan LSTM utama, di mana lapisan pertama (lstm_681) memiliki 64 unit dengan konfigurasi *return sequences* aktif, diikuti oleh lapisan *Dropout* dengan laju 0.2 untuk meminimalisir risiko *overfitting*. Selanjutnya, lapisan LSTM kedua (lstm_682) dengan 64 unit digunakan untuk mengekstraksi fitur sekuensial secara mendalam sebelum diteruskan ke lapisan *Dense* sebagai *output* tunggal. Hasil visualisasi pada siklus 24 jam menunjukkan bahwa model mampu mengikuti tren fluktuasi harian polutan di berbagai wilayah pemantauan, meskipun terdapat deviasi pada titik-titik puncak (*peak*) konsentrasi tinggi. Penggunaan pendekatan *multivariate* dengan melibatkan seluruh wilayah serta penambahan fitur temporal jam terbukti meningkatkan stabilitas model dalam menghasilkan prediksi nilai kontinu konsentrasi polutan udara secara akurat.

```
Processing region: Denpasar
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 49ms/step
Finished Denpasar. RMSE: 2.21, R2: 0.743
------------------------------------------------------------
Processing region: Badung
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 31ms/step
Finished Badung. RMSE: 4.24, R2: 0.839
------------------------------------------------------------
Processing region: Gianyar
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 30ms/step
Finished Gianyar. RMSE: 4.15, R2: 0.846
------------------------------------------------------------
Processing region: Tabanan
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 31ms/step
Finished Tabanan. RMSE: 4.36, R2: 0.830
------------------------------------------------------------
Processing region: Bangli
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 40ms/step
Finished Bangli. RMSE: 4.16, R2: 0.845
------------------------------------------------------------
Processing region: Klungkung
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 30ms/step
Finished Klungkung. RMSE: 1.83, R2: 0.853
------------------------------------------------------------
Processing region: Karangasem
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 31ms/step
Finished Karangasem. RMSE: 1.84, R2: 0.850
------------------------------------------------------------
Processing region: Singaraja
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 30ms/step
Finished Singaraja. RMSE: 2.51, R2: 0.827
------------------------------------------------------------
Processing region: Jembrana
22/22 ━━━━━━━━━━━━━━━━━━━━ 1s 31ms/step
Finished Jembrana. RMSE: 2.35, R2: 0.711
------------------------------------------------------------
Semua wilayah selesai diproses.
```

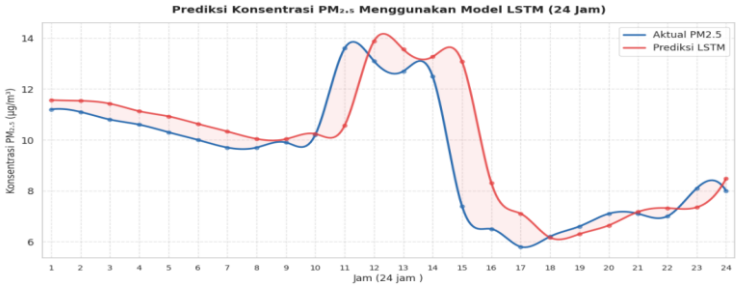
Gambar 7. Ringkasan Kinerja Model LSTM per Wilayah.

Gambar tersebut menyajikan ringkasan kinerja model LSTM standar pada masing-masing wilayah pengukuran di Bali dan sekitarnya, yang dievaluasi menggunakan dua metrik utama: Root Mean Square Error (RMSE) dan koefisien determinasi (R^2 Score). Hasil menunjukkan variasi performa yang cukup signifikan antar lokasi. Hasil evaluasi model Long Short-Term Memory (LSTM) menunjukkan performa yang bervariasi antar wilayah di Bali. Model LSTM mencapai nilai Root Mean Squared Error (RMSE) terendah sebesar 1,83 pada wilayah Klungkung, diikuti oleh Karangasem dengan RMSE 1,84. Nilai RMSE tertinggi tercatat di wilayah Badung yaitu 4,24, sementara Denpasar sebagai ibu kota provinsi memperoleh RMSE sebesar 2,21. Koefisien determinasi (R^2) tertinggi diperoleh di wilayah Gianyar sebesar 0,846, yang menunjukkan model mampu menjelaskan variasi data PM2.5 cukup baik. Secara keseluruhan, nilai R^2 pada semua wilayah berada di atas 0,71, mengindikasikan bahwa model LSTM memiliki kemampuan prediksi yang cukup baik. Proses komputasi model LSTM relatif cepat dengan waktu pemrosesan rata-rata sekitar 30–40 detik per wilayah. Hasil ini menunjukkan bahwa model LSTM dapat diterapkan secara efektif untuk memprediksi kadar PM2.5 di berbagai wilayah di Bali dengan tingkat akurasi yang memadai.

Evaluasi dan Analisis Akhir

Tahap evaluasi dan analisis akhir bertujuan untuk mengukur performa model LSTM dalam menghasilkan prediksi konsentrasi $PM_{2.5}$ yang akurat dan reliabel di seluruh wilayah

pemantauan. Pengujian dilakukan dengan membandingkan nilai prediksi model terhadap data aktual pada *test set* menggunakan metrik statistik standar, yaitu *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan Koefisien Determinasi R^2. Analisis ini tidak hanya berfokus pada akurasi numerik, tetapi juga mengevaluasi kemampuan model dalam menangkap pola temporal dan fluktuasi puncak polutan yang terjadi pada siklus harian. Dengan melibatkan seluruh fitur wilayah (*multivariate*) dan fitur temporal jam, hasil evaluasi ini akan memberikan gambaran komprehensif mengenai stabilitas model serta sejauh mana generalisasi model dapat diterapkan pada kondisi kualitas udara yang dinamis.



Gambar 8. Grafik Model LSTM.

Gambar 8 Model Long Short-Term Memory (LSTM) yang dikembangkan dalam penelitian ini menunjukkan performa yang baik dalam memprediksi kadar Particulate Matter 2.5 (PM2.5) secara waktu nyata selama 24 jam. Sebagaimana terlihat pada Gambar prediksi akhir, garis prediksi LSTM (merah) sangat mengikuti pola fluktuasi data aktual (biru), terutama pada peningkatan tajam konsentrasi antara jam 10–14 dan penurunan drastis setelah jam 15. Nilai Root Mean Squared Error (RMSE) dan koefisien determinasi (R^2) yang diperoleh pada sembilan wilayah di Bali berkisar antara 1,83–4,36 dan 0,711–0,853 secara berturut-turut, membuktikan bahwa model mampu menjelaskan lebih dari 71 % variasi data PM2.5 dengan tingkat kesalahan yang rendah. Area bayangan merah pada grafik juga menunjukkan selisih prediksi yang relatif kecil, terutama pada periode stabil (jam 1–9 dan 18–24). Keberhasilan LSTM dalam menangkap pola temporal jangka panjang ini menegaskan keunggulannya dibandingkan model statistik tradisional seperti ARIMA.

Secara keseluruhan, hasil evaluasi akhir membuktikan bahwa model LSTM dapat diandalkan sebagai sistem prediksi dini kualitas udara di wilayah Bali. Meskipun terdapat sedikit overestimation pada puncak konsentrasi siang hari dan underestimation pada malam hari, deviasi tersebut masih berada dalam batas toleransi untuk aplikasi pemantauan lingkungan. Temuan ini memberikan kontribusi praktis bagi pemerintah daerah dan BMKG dalam menyusun peringatan dini paparan PM2.5, khususnya bagi kelompok rentan di kota-kota besar seperti Denpasar dan Badung. Penelitian ini membuka peluang pengembangan lebih

lanjut, seperti integrasi LSTM dengan data satelit atau optimasi hyperparameter menggunakan algoritma metaheuristik untuk meningkatkan akurasi prediksi jangka menengah. Dengan demikian, model LSTM terbukti efektif dan layak digunakan sebagai pendekatan deep learning berbasis time series untuk mitigasi pencemaran udara di Indonesia.

Pembahasan

Penelitian ini membuktikan bahwa model Long Short-Term Memory (LSTM) menunjukkan kemampuan yang baik dalam menangkap pola temporal fluktuasi kadar PM2.5 selama 24 jam. Grafik prediksi akhir menggambarkan bahwa garis prediksi LSTM (merah) cukup mengikuti pola data aktual (biru), terutama pada kenaikan tajam konsentrasi antara pukul 10 hingga 14 serta penurunan signifikan setelah pukul 15. Area bayangan merah yang relatif sempit mengindikasikan selisih prediksi yang kecil pada sebagian besar periode, meskipun terdapat sedikit overestimation pada puncak siang hari dan underestimation pada malam hari. Kemampuan LSTM dalam mengingat informasi jangka panjang melalui mekanisme gate memungkinkan model ini menangani ketergantungan sekuensial yang kompleks pada data PM2.5 yang dipengaruhi faktor meteorologi dan aktivitas antropogenik. Secara keseluruhan, visualisasi ini membuktikan bahwa arsitektur LSTM efektif untuk memodelkan dinamika harian polutan halus di wilayah tropis seperti Bali.

Performa model LSTM bervariasi antar wilayah, dengan nilai RMSE terendah tercatat di Klungkung (1,83) dan Karangasem (1,84), serta R^2 tertinggi di Gianyar (0,846) dan Klungkung (0,853). Hal ini menunjukkan bahwa model lebih akurat pada wilayah dengan tingkat fluktuasi PM2.5 yang relatif stabil dibandingkan wilayah dengan aktivitas lalu lintas dan industri yang lebih tinggi seperti Badung (RMSE 4,24). Nilai R^2 di atas 0,71 pada semua wilayah mengindikasikan bahwa model mampu menjelaskan lebih dari 71% variasi data aktual, yang masih sejalan dengan berbagai penelitian sebelumnya yang melaporkan R^2 LSTM untuk prediksi PM2.5 berkisar antara 0,52 hingga 0,91. Perbedaan performa antar wilayah kemungkinan disebabkan oleh karakteristik lokal, seperti kepadatan penduduk, topografi, dan pola angin yang memengaruhi dispersi polutan. Hasil ini menegaskan pentingnya pengembangan model berbasis lokasi (location-specific model) untuk meningkatkan akurasi prediksi.

Meskipun menunjukkan hasil yang memuaskan, model LSTM masih memiliki keterbatasan, terutama pada prediksi puncak ekstrem dan periode dengan perubahan mendadak akibat faktor eksternal seperti kebakaran lahan atau kondisi cuaca ekstrem. Deviasi kecil yang terlihat pada grafik prediksi dapat diminimalkan melalui penambahan fitur eksternal, integrasi model hybrid (misalnya CNN-LSTM), atau optimasi hyperparameter lebih lanjut. Temuan

penelitian ini memberikan kontribusi praktis bagi sistem pemantauan kualitas udara di Bali, khususnya dalam mendukung peringatan dini bagi masyarakat rentan. Ke depannya, penelitian dapat dikembangkan dengan data real-time yang lebih panjang serta penggabungan algoritma metaheuristik untuk meningkatkan generalisasi model di berbagai wilayah Indonesia.

4. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan model Long Short-Term Memory (LSTM) untuk memprediksi kadar Particulate Matter 2.5 (PM2.5) di sembilan wilayah di Bali. Model LSTM menunjukkan performa yang cukup baik dengan nilai RMSE berkisar antara 1,83 hingga 4,36 dan koefisien determinasi (R^2) antara 0,711 hingga 0,853. Visualisasi kurva prediksi 24 jam mengonfirmasi bahwa garis prediksi LSTM mampu mengikuti pola fluktuasi data aktual dengan selisih yang relatif kecil, terutama pada periode peningkatan siang hari dan penurunan malam hari. Hasil evaluasi membuktikan bahwa arsitektur LSTM efektif menangkap ketergantungan temporal jangka panjang pada data time series PM2.5 yang dipengaruhi oleh faktor meteorologi dan aktivitas antropogenik. Secara keseluruhan, model ini telah memenuhi tujuan penelitian sebagai salah satu pendekatan deep learning yang andal untuk prediksi kualitas udara. Meskipun hasil yang diperoleh positif, penelitian ini masih memiliki keterbatasan berupa penggunaan data historis yang terbatas dan belum memasukkan variabel eksternal seperti data satelit atau kejadian kebakaran lahan. Perbedaan performa antar wilayah menunjukkan pentingnya pengembangan model berbasis lokasi. Untuk penelitian selanjutnya, disarankan melakukan integrasi model hybrid (CNN-LSTM), optimasi hyperparameter menggunakan algoritma metaheuristik, serta pengembangan sistem prediksi real-time. Temuan penelitian ini diharapkan dapat menjadi kontribusi bagi pemerintah daerah Bali dalam menyusun strategi mitigasi pencemaran udara dan memberikan peringatan dini bagi masyarakat rentan.

DAFTAR PUSTAKA

Bai, X., Zhang, N., Cao, X., & Chen, W. (2024). Prediction of PM2.5 concentration based on a CNN-LSTM neural network algorithm. *PeerJ, 12*, e17811. https://doi.org/10.7717/peerj.17811

Basith, S., et al. (2022). The impact of fine particulate matter 2.5 on the cardiovascular system: A review of the invisible killer. *Nanomaterials, 12*(15), 2656. https://doi.org/10.3390/nano12152656

Guo, Z., Wu, S., Zhu, M., & He, G. (2025). Air quality PM2.5 index prediction model based on CNN-LSTM.

Hadi, N., Alam, S., & Kurniawan, I. (2025). Penerapan model LSTM (Long Short-Term Memory) dalam peramalan inflasi month-on-month di Indonesia. *JATI: Jurnal Mahasiswa Teknik Informatika*. https://doi.org/10.36040/jati.v9i5.15119

Hasibuan, M. (2025). Student enrollment forecasting system using ARIMA time series prediction. *Jurnal Provisi*. Retrieved from https://ejurnal.provisi.ac.id/index.php/JUISI/article/view/1317

Ho, C.-H., Park, I., Kim, J., & Lee, J.-B. (2023). PM2.5 forecast in Korea using the Long Short-Term Memory (LSTM) model. *Asia-Pacific Journal of Atmospheric Sciences*, *59*(5), 563-576. https://doi.org/10.1007/s13143-022-00293-2

Karyadi, Y. (2022). Prediksi kualitas udara dengan metoda LSTM, bidirectional LSTM, dan GRU. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi*, *9*(1), 671-684. https://doi.org/10.35957/jatisi.v9i1.1588

Khanthavit, R., & Khanthavit, A. (2025). Particulate matter 2.5 pollution, perception, and mental stress. *International Journal of Public Health Science (IJPHS)*, *14*(1), 292. https://doi.org/10.11591/ijphs.v14i1.24587

Kristiani, E., Lin, H., Lin, J.-R., Chuang, Y.-H., Huang, C.-Y., & Yang, C.-T. (2022). Short-term prediction of PM2.5 using LSTM deep learning methods. *Sustainability*, *14*(4), 2068. https://doi.org/10.3390/su14042068

Meng, X., Xie, C., Tang, X., & Pan, Y. (2025). Prediction of particulate matter 2.5 concentration based on attention mechanism and convolutional BiLSTM network. *Discover Applied Sciences*, *7*(11), 1372. https://doi.org/10.1007/s42452-025-07891-5

Nourmohammad, E., & Rashidi, Y. (2025). Ground data analysis for PM2.5 prediction using predictive modeling techniques. *Journal of Air Pollution and Health*. https://doi.org/10.18502/japh.v10i1.18095

Sitompul, P. A. U., Sujaini, H., Yacoub, R. R., & Ananda, N. (2025). Penerapan model Long Short-Term Memory (LSTM) untuk prediksi kualitas udara dan analisis korelasi terhadap unsur cuaca di Jongkat Kabupaten Mempawah.

Wahyutama, S., Rismawan, T., & Bahri, S. (2026). Penerapan algoritma Long Short Term Memory (LSTM) dalam prediksi pergerakan harga cryptocurrency menggunakan data time series. *Universitas Tanjungpura*. https://doi.org/10.26418/coding.v14i1.98866

Wang, J., Han, J., Li, T., Wu, T., & Fang, C. (2023). Impact analysis of meteorological variables on PM2.5 pollution in the most polluted cities in China. *Heliyon*, *9*(7), e17609. https://doi.org/10.1016/j.heliyon.2023.e17609

Yang, R. (2025). Time series analysis: Application of LSTM model in predicting PM2.5 concentration in Beijing. *ITM Web of Conferences*, *70*, 04022. https://doi.org/10.1051/itmconf/20257004022