

IMPUTASI MISSING VALUE DATA TEMPERATUR UDARA MENGGUNAKAN METODE MULTILAYER PERCEPTRON (MLP)

Nurul Faiza Ie^{1*}, Muchammad Rizqy Nugraha¹, Marzuki Sinambela¹, Edward Trihadi¹, Dyah Prihartini Djenal²

¹Instrumentasi-MKG, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika Jalan Meteorologi Nomor 5, Kecamatan Tanah Tinggi, Kota Tangerang, Provinsi Banten

²Badan Meteorologi Klimatologi dan Geofisika Jl. Angkasa I No.2 Kemayoran, Jakarta Pusat, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta

Keywords:

Imputasi missing value
Temperatur udara
Multilayer Perceptron
Data meteorologi
Machine learning

Correspondent Email:

nrulfaiza@stmkg.ac.id

Abstrak. Kelengkapan dan akurasi data temperatur udara merupakan kebutuhan utama dalam studi meteorologi serta berbagai aplikasi berbasis data cuaca. Namun pada kenyataannya, data pengukuran dari sensor sering mengalami kekosongan nilai akibat kendala teknis pada alat maupun pengaruh faktor eksternal. Penelitian ini mengembangkan model imputasi untuk menutup data temperatur yang hilang dengan menggunakan metode Multilayer Perceptron (MLP). Fitur yang digunakan meliputi kelembaban, tekanan udara, radiasi matahari, dan temperatur pada waktu sebelumnya sebagai variabel lag. Dataset yang digunakan terdiri dari 41.410 baris data historis yang telah melalui proses pembersihan. Pembagian data dilakukan dengan rasio 80% untuk pelatihan dan 20% untuk pengujian secara kronologis (*time-based*), serta dilakukan normalisasi menggunakan StandardScaler. Hasil evaluasi menunjukkan model MLP dengan dua *hidden layer* berukuran 150 dan 75 neuron mampu memprediksi temperatur dengan baik. Hal ini dibuktikan dengan nilai RMSE 0,868°C, MAE 0,591°C, dan R² 0,903. Hasil ini menunjukkan bahwa pendekatan MLP dengan fitur temporal efektif digunakan untuk imputasi data temperatur yang hilang.



Copyright © 2026 The Author(s),
Published by Jurnal Informatika
dan Teknik Elektro Terapan. This
is an open-access article
distributed under the terms of the
Creative Commons Attribution-
NonCommercial 4.0 International
License (CC BY-NC 4.0).

Abstract. Complete and accurate air temperature data is essential for meteorological analysis and other weather-based applications. In practice, however, sensor measurements often contain missing values due to instrument malfunctions or external disturbances. This study proposes a Multilayer Perceptron (MLP) approach to impute missing air temperature values. The model utilizes four predictor variables: humidity, air pressure, solar radiation, and the temperature from the previous time step as a lag feature. A total of 41,410 historical records were used after undergoing a data cleaning process. The dataset was split chronologically (*time-based*) into 80% for training and 20% for testing, with data normalized using StandardScaler. The evaluation results show that an MLP model with two hidden layers of 150 and 75 neurons achieved high prediction accuracy, indicated by an RMSE of 0.868°C, an MAE of 0.591°C, and an R² of 0.903. These findings indicate that the MLP approach incorporating temporal features is effective for imputing missing air temperature data.

1. PENDAHULUAN

Data meteorologi merupakan unsur krusial dalam proses pengambilan keputusan di

berbagai bidang, seperti prakiraan cuaca, studi iklim, sektor pertanian, hingga perencanaan sistem energi[1]. Dari seluruh parameter yang diukur, temperatur udara termasuk variabel dasar yang memiliki dampak besar terhadap kondisi atmosfer maupun aktivitas manusia. Oleh karena itu, kelengkapan dan kesinambungan data temperatur menjadi prasyarat utama agar hasil analisis memiliki akuntabilitas ilmiah[2].

Dalam pelaksanaannya, pengukuran temperatur menggunakan sensor otomatis atau *Automatic Weather Station* sering kali tidak berjalan optimal. Gangguan pada sumber daya listrik, kerusakan komponen sensor, kegagalan dalam transmisi data, serta kegiatan kalibrasi ulang alat dapat menyebabkan munculnya data kosong atau *missing value*[3][4]. Jika kondisi tersebut tidak ditangani, maka akan muncul bias dalam analisis statistik dan penurunan akurasi model prediksi yang disusun berdasarkan data tersebut. Oleh karena itu, diperlukan metode imputasi yang dapat mengestimasi nilai yang hilang secara tepat dengan memanfaatkan keterkaitan antar parameter meteorologi yang datanya tersedia lengkap[5].

Beragam teknik imputasi telah dikembangkan untuk menangani persoalan data hilang, mulai dari metode statistik dasar seperti imputasi rata-rata dan interpolasi linear, hingga model regresi linear berganda[6]. Akan tetapi, pendekatan-pendekatan tersebut masih memiliki kelemahan dalam merepresentasikan hubungan non-linear antar variabel meteorologi yang secara alami bersifat kompleks[7]. Sebagai alternatif, metode *machine learning* terutama *Artificial Neural Network* (ANN) dengan arsitektur Multilayer Perceptron (MLP) dinilai lebih unggul karena mampu memodelkan keterkaitan non-linear antara variabel prediktor dan variabel target dengan fleksibilitas yang lebih tinggi[8].

Beberapa studi terdahulu telah menerapkan pendekatan *machine learning* untuk imputasi data meteorologi dengan menggunakan sumber data yang sama, tetapi pada periode waktu berbeda[7]. Hal tersebut memberikan celah bagi penelitian ini untuk melakukan studi replikasi sekaligus menguji konsistensi kinerja model MLP pada rentang data yang baru. Selain itu, penelitian ini juga mengeksplorasi penambahan fitur temporal berupa nilai temperatur pada

waktu sebelumnya (*lag feature*) untuk meningkatkan akurasi estimasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model MLP dalam mengimputasi data temperatur udara yang hilang dengan memanfaatkan kelembaban udara, tekanan udara, radiasi matahari, dan temperatur periode sebelumnya sebagai prediktor serta mengevaluasi performa model menggunakan metrik RMSE, MAE, dan R^2 dengan pembagian data latih dan uji 80:20 secara kronologis.

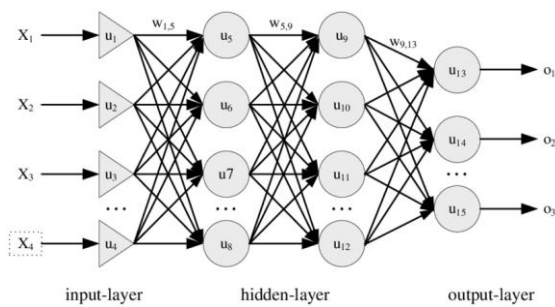
2. TINJAUAN PUSTAKA

2.1 *Missing Value Pada Data Meteorologi*

Data yang hilang atau *missing value* merujuk pada tidak adanya nilai pada satu atau lebih variabel dalam dataset[9]. Pada data meteorologi berbasis deret waktu, fenomena ini biasanya bersifat acak dalam rentang waktu tertentu dan dipicu oleh gangguan pada alat ukur. Penanganan yang keliru terhadap data hilang dapat mengakibatkan berkurangnya informasi penting dan munculnya bias dalam analisis statistik.

2.2 *Multilayer Perceptron*

Multilayer Perceptron (MLP) adalah jenis arsitektur *feedforward neural network* yang tersusun atas lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output[10]. Pada lapisan tersembunyi, setiap neuron melakukan transformasi non-linear terhadap hasil kombinasi linear dari input melalui fungsi aktivasi, contohnya *Rectified Linear Unit* (ReLU)[11]. Tahap pelatihan pada MLP menggunakan algoritma *backpropagation* untuk meminimalkan fungsi kerugian dengan bantuan optimizer seperti Adam (*Adaptive Moment Estimation*). Fleksibilitas MLP dalam merepresentasikan hubungan non-linear antar variabel menjadikannya metode yang sesuai untuk menyelesaikan permasalahan regresi pada data meteorologi yang memiliki keterkaitan antar parameter yang kompleks[8].



Gambar 1. Arsitektur Multilayer Perceptron (MLP)

2.3 Hubungan Antar Variabel Meteorologi

Dari aspek fisika, temperatur udara memiliki keterkaitan dengan berbagai parameter meteorologi lainnya. Kelembaban udara cenderung berbanding terbalik dengan temperatur, sementara tekanan udara menunjukkan hubungan yang kompleks karena dipengaruhi oleh faktor ketinggian dan kondisi atmosfer[12][13]. Radiasi matahari juga berperan secara langsung dalam meningkatkan temperatur permukaan, khususnya pada periode siang[14]. Di samping keterkaitan antar variabel tersebut, temperatur udara memiliki sifat autokorelasi temporal yang kuat. Hal ini berarti nilai temperatur pada suatu waktu tertentu dipengaruhi oleh nilai temperatur pada waktu sebelumnya, seiring dengan perubahan atmosfer yang berlangsung secara bertahap.

2.4 Penelitian Terdahulu

Studi tentang prediksi dan imputasi parameter meteorologi berbasis *machine learning* dan deep learning telah banyak dilakukan sebelumnya, baik menggunakan sumber data dari BMKG maupun pada lokasi pengamatan yang berbeda. Salah satunya adalah penelitian oleh Suciarti dkk, (2023) yang fokus pada estimasi kekosongan data temperatur udara dari sensor *Automatic Weather Station* (AWS). Penelitian tersebut mengambil studi kasus di AWS Cisurupan, Kabupaten Garut, Jawa Barat, dengan data beresolusi 10 menit tahun 2019. Hasilnya menunjukkan bahwa model hibrida ARIMA-MLP efektif dalam menangkap karakteristik linear dan non-linear pada data temperatur. Penelitian tersebut juga menggunakan proporsi data latih minimal 85% sesuai rekomendasi WMO dan memperoleh nilai RMSE di bawah 0,20°C. Secara metodologi, studi ini memiliki

relevansi dengan penelitian saat ini, terutama terkait pemanfaatan MLP untuk menangani *missing value* pada data temperatur dari sensor AWS[15].

Penelitian terbaru oleh Panji dan Sofwan (2025) mengembangkan tiga model ANN berbasis algoritma *backpropagation* untuk melakukan prediksi simultan terhadap suhu udara, kelembaban relatif, dan tekanan udara. Data yang digunakan merupakan data satu menit selama periode Januari–Desember 2024 dari Stasiun Meteorologi Fatmawati Soekarno, Bengkulu, dengan pembagian data latih, validasi, dan uji sebesar 80:10:10. Hasilnya menunjukkan model prediksi suhu udara dengan arsitektur dua *hidden layer* masing-masing 40 dan 20 neuron menghasilkan kinerja terbaik, ditandai dengan nilai koefisien korelasi R sebesar 0,915 dan RMSE 1,198°C. Sebaliknya, prediksi tekanan udara memiliki korelasi yang rendah dengan R = 0,446, yang diduga karena pengaruh faktor atmosfer skala besar yang tidak tercakup dalam variabel input lokal. Temuan tersebut mendukung pemilihan variabel prediktor pada penelitian ini dan sekaligus menjadi rujukan metodologis, mengingat kesamaan sumber data dari BMKG meskipun berbeda periode dan lokasi pengamatan[16].

Studi oleh Supriyadi, (2020) mengimplementasikan model LSTM untuk prediksi parameter cuaca jangka pendek berbasis data BMKG dua bulan pertama tahun 2019. Fitur yang digunakan mencakup suhu, kelembaban, angin, dan tekanan udara dengan pembagian data 9:1. Model dilengkapi pembaruan berkala untuk mempertahankan performa pada horizon prediksi yang lebih panjang. Evaluasi menunjukkan RMSE suhu 0,576°C, menandakan performa yang sangat baik. Akan tetapi, kesalahan prediksi suhu dan kelembaban meningkat pada horizon yang lebih jauh, sementara angin dan tekanan menunjukkan stabilitas. Temuan ini mendukung relevansi penelitian saat ini yang juga memanfaatkan dependensi temporal melalui fitur `_temp_lag1_` untuk meningkatkan akurasi imputasi suhu[17].

Tiga studi terdahulu secara konsisten menunjukkan bahwa metode *machine learning* dan deep learning mampu memodelkan hubungan non-linear dan temporal antar variabel meteorologi dengan akurasi tinggi,

terutama pada temperatur udara. Akan tetapi, penelitian Suciarti, dkk (2023) dan Panji dan Sofwan (2025) belum menguji penggunaan fitur temporal berupa temperatur periode sebelumnya sebagai input MLP. Sementara itu, Supriyadi (2020) menggunakan arsitektur LSTM yang memang dirancang untuk data deret waktu, namun memiliki kompleksitas komputasi lebih tinggi dan mengalami penurunan akurasi pada horizon prediksi yang panjang [15][16][17]. Berbeda dengan itu, penelitian ini mengusulkan pendekatan yang lebih sederhana dengan menambahkan fitur `_temp_lag1_` ke dalam MLP

2.5 Metrik Evaluasi

Evaluasi model dilakukan menggunakan tiga metrik evaluasi yaitu MAE, RMSE, dan R^2 , sebagaimana dirumuskan pada Persamaan (1) – (3)[18].

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

3. METODE PENELITIAN

3.1 Data Penelitian

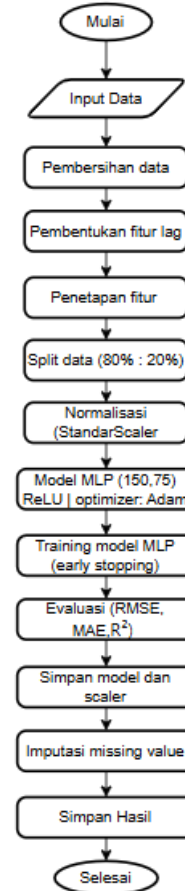
Dataset yang digunakan dalam penelitian ini berupa data suhu udara per jam yang tercatat oleh AWS di Stasiun Meteorologi Kelas I Soekarno-Hatta. Rentang waktu data mencakup periode 1 Januari 2020 sampai dengan 31 Desember 2024. Data historis tersebut diperoleh dari Kantor Pusat BMKG melalui pengajuan permohonan resmi dan sebelumnya telah melalui tahapan kendali mutu (*quality control*). Data terdiri atas lima variabel utama, yaitu:

Tabel 1. Variabel Utama yang Digunakan

Variabel	Keterangan
Kelembaban udara	Variabel prediktor
Tekanan udara	Variabel prediktor
Radiasi matahari	Variabel prediktor
Temp_lag1	Variabel prediktor
Temperatur udara	Variabel target

3.2 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahap sistematis sebagaimana ditunjukkan pada diagram alir berikut.



Gambar 2. Diagram Alir Penelitian

Gambar 2 adalah diagram alir dari penelitian imputasi data hilang menggunakan model MLP secara umum. Penjelasan dari diagram alir tersebut adalah sebagai berikut:

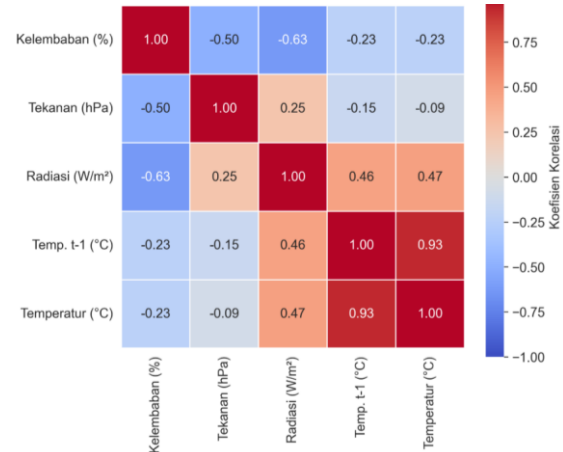
1. Mulai : Titik awal dari seluruh rangkaian proses penelitian.
2. Pengumpulan Data Mentah : Memperoleh raw data historis meteorologi yang terdiri dari variabel kelembaban, tekanan, radiasi, dan temperatur.
3. Pembersihan Data : Melakukan deteksi dan penanganan *outlier* pada variabel prediktor dan target, serta penanganan awal *missing value* pada fitur. Tahap ini menghasilkan dataset bersih sebanyak 41.410 baris dan dilakukan terpisah sebelum pemodelan.
4. Pembentukan Fitur Lag : Membentuk fitur `_temp_lag1_` dari data yang sudah bersih

- untuk merepresentasikan autokorelasi temporal temperatur.
5. Penetapan Fitur: Menetapkan variabel input berupa kelembaban, tekanan, radiasi, dan `_temp_lag1_`.
 6. Pembagian Data : Membagi dataset menjadi data latih dan data uji dengan rasio 80% : 20% secara kronologis.
 7. Normalisasi Data : Melakukan penskalaan pada fitur dan target menggunakan StandardScaler. Proses fitting hanya dilakukan pada data latih untuk mencegah *data leakage*.
 8. Perancangan Arsitektur MLP : Membangun model MLP dengan 2 *hidden layer* berukuran 150 dan 75 neuron, menggunakan fungsi aktivasi ReLU dan optimizer Adam.
 9. Pelatihan Model : Melatih model dengan menerapkan *mekanisme early stopping* untuk mencegah terjadinya *overfitting*.
 10. Evaluasi Model : Menguji performa model pada data uji menggunakan metrik RMSE, MAE, dan R^2 .
 11. Penyimpanan Model dan Scaler : Menyimpan model dan scaler hasil pelatihan dalam format .pkl untuk digunakan kembali.
 12. Imputasi *Missing Value* Temperatur : Menerapkan model yang telah dilatih untuk mengisi *missing value* pada kolom temperatur. Tahap ini berbeda dengan pembersihan awal yang hanya menangani *missing value* pada fitur prediktor.
 13. Simpan hasil (.csv): Data akhir yang sudah terisi disimpan sebagai output.
 14. Selesai: akhir proses

4. HASIL DAN PEMBAHASAN

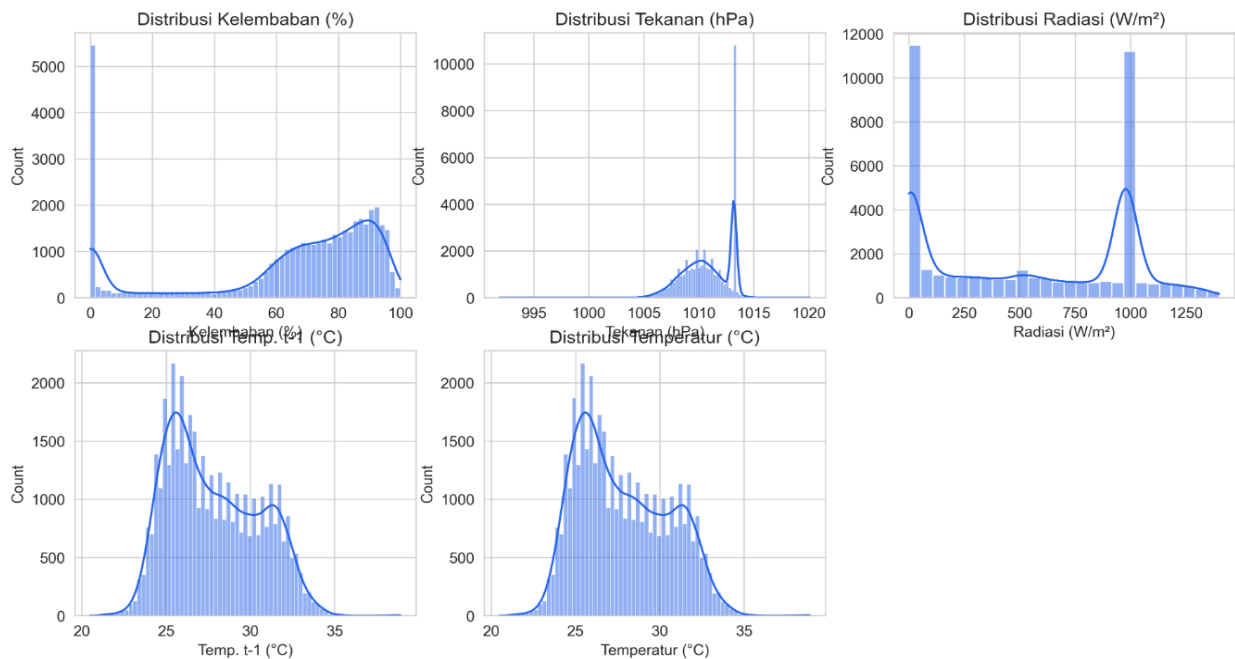
4.1 Eksplorasi dan Karakteristik Data

Sebelum pemodelan dilakukan, perlu dilakukan eksplorasi data terlebih dahulu untuk memahami karakteristik dan hubungan antar variabel.



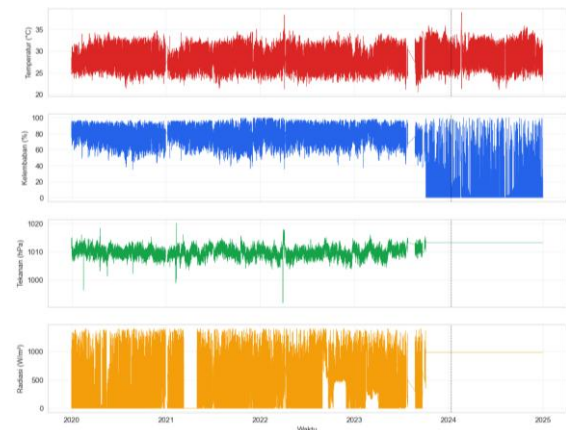
Gambar 3. Matriks Korelasi Antar Variabel

Hasil analisis korelasi menunjukkan bahwa temperatur udara memiliki hubungan paling kuat dengan `_temp_lag1_` dengan nilai $r = 0,93$, dan diikuti oleh radiasi matahari dengan $r = 0,47$. Sebaliknya, kelembaban memiliki korelasi negatif terhadap temperatur sebesar $r = -0,23$, sementara tekanan udara menunjukkan korelasi yang sangat lemah dengan $r = -0,09$. Selain itu, kelembaban juga berkorelasi negatif cukup kuat dengan radiasi matahari $r = -0,63$ dan dengan tekanan $r = -0,50$. Hal ini sejalan dengan prinsip fisis atmosfer, di mana kondisi lembab umumnya terjadi pada saat intensitas radiasi matahari rendah. Temuan ini menegaskan bahwa komponen temporal melalui `_temp_lag1_` memiliki pengaruh yang jauh lebih dominan dibandingkan variabel meteorologi sesaat, sehingga menjadi landasan penggunaan fitur `_temp_lag1_` dalam model penelitian ini.



Gambar 4. Distribusi Variabel

Distribusi variabel kelembaban menunjukkan karakteristik bimodal. Terdapat lonjakan nilai yang tajam di sekitar 0% dan distribusi utama yang terkonsentrasi pada rentang 60–100%. Untuk tekanan udara, sebagian besar data terpusat pada nilai 1013 hPa dengan frekuensi kemunculan yang tinggi, disertai sebaran sekunder yang lebih landai pada rentang 1005–1015 hPa. Variabel radiasi matahari juga memiliki pola bimodal, ditandai oleh puncak dominan di sekitar 0 W/m² yang merepresentasikan kondisi malam atau tanpa radiasi, serta puncak kedua di sekitar 1000 W/m² yang merepresentasikan kondisi siang dengan intensitas radiasi tinggi. Sementara itu, temperatur udara dan `_temp_lag1_` memiliki pola distribusi yang hampir sama, yaitu mendekati distribusi normal dengan puncak utama pada kisaran 25–26°C dan gelombang kecil di rentang 30–32°C. Kemiripan ini mengonfirmasi adanya keterkaitan karakteristik antara kedua variabel akibat pengaruh lag satu periode waktu.

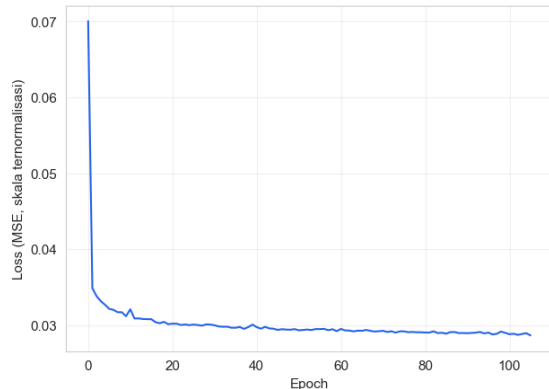


Gambar 5. Deret Waktu Seluruh Variabel

Keempat variabel memperlihatkan pola fluktuasi harian yang konsisten selama periode 2020–2023. Hal ini termasuk adanya pola musiman pada radiasi matahari yang mengikuti siklus siang-malam, serta pada temperatur udara. Akan tetapi, mulai akhir tahun 2023 teridentifikasi perubahan karakteristik data yang signifikan. Nilai tekanan udara dan radiasi matahari berubah menjadi konstan, masing-masing berada di sekitar 1013 hPa dan 1000 W/m². Di sisi lain, kelembaban menunjukkan variabilitas yang lebih tinggi dengan kemunculan nilai rendah yang lebih sering. Pola tersebut mengindikasikan adanya kemungkinan pergantian sumber data, gangguan pada sensor, atau perbedaan prosedur pencatatan pada periode tersebut. Oleh karena itu, temuan ini

perlu dijadikan catatan penting dalam interpretasi hasil dan sebagai salah satu keterbatasan penelitian. Perlu dicatat bahwa periode anomali ini tumpang tindih dengan bagian akhir data latih dan hampir keseluruhan data uji, sehingga variasi prediktor tekanan dan radiasi pada periode evaluasi menjadi sangat terbatas.

4.2 Pelatihan Model



Gambar 6. Kurva Loss Training MLP

Kurva loss mengalami penurunan yang signifikan pada fase awal pelatihan, dari nilai sekitar 0,070 menjadi 0,032 dalam beberapa belas epoch pertama. Setelah fase tersebut, laju penurunan melambat tetapi tetap stabil, dengan nilai loss berangsur turun hingga mencapai 0,029 di akhir pelatihan. Terdapat fluktuasi kecil di sekitar epoch ke-10, tetapi tidak mengubah pola penurunan secara umum. Pelatihan dihentikan otomatis pada epoch ke-106 melalui mekanisme early stopping, yang menunjukkan bahwa model telah mencapai konvergensi. Tidak ditemukan indikasi *underfitting* maupun *overfitting* pada tahap akhir.

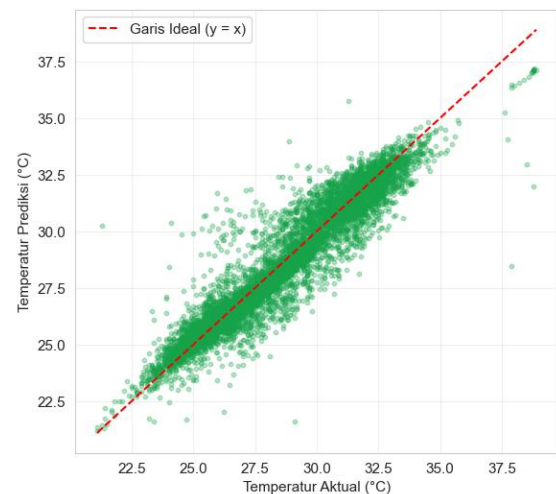
4.3 Evaluasi Performa Model

Tabel 2. Evaluasi Performa Model

Metrik	Nilai
MAE	0,591°C
RMSE	0,868°C
R ²	0,903

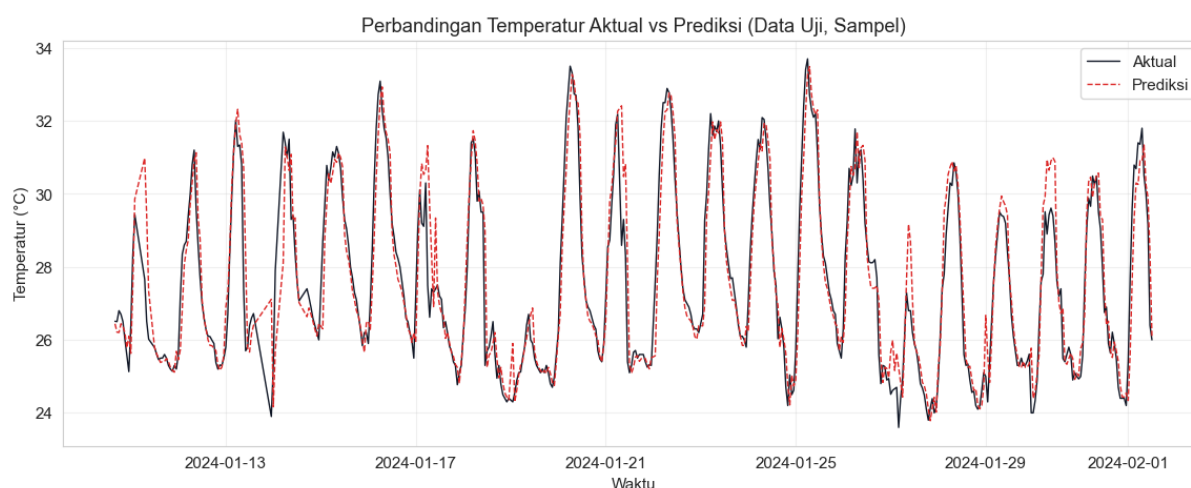
Performa model pada data uji menunjukkan nilai R² sebesar 0,90. Nilai tersebut mengindikasikan bahwa model dapat menjelaskan 90% dari total variansi temperatur udara. Dengan demikian, model memiliki

tingkat akurasi yang tinggi dalam merepresentasikan pola data. Sementara itu, nilai MAE sebesar 0,591°C mengartikan bahwa rata-rata selisih antara hasil prediksi model dan data aktual berada di bawah 0,6°C. Adapun nilai RMSE yang tercatat lebih tinggi yaitu 0,868°C, mengindikasikan keberadaan sejumlah prediksi dengan deviasi yang cukup besar. Hal ini terjadi karena RMSE memberikan penalti lebih tinggi terhadap kesalahan besar atau *outlier* dibandingkan dengan MAE.



Gambar 7. Scatter Plot Antara Nilai Aktual dan Prediksi

Scatter plot antara nilai aktual dan prediksi memperlihatkan bahwa mayoritas data pada kisaran temperatur 20–35°C tersebar berdekatan dengan garis referensi $y = x$. Hal ini menandakan bahwa model memiliki tingkat akurasi yang tinggi dalam memprediksi temperatur pada kondisi normal. Sebaliknya, pada rentang temperatur tinggi sekitar 38–40°C, terdapat kumpulan titik yang secara konsisten terletak di bawah garis diagonal, yang mengindikasikan adanya kecenderungan *underestimate* oleh model pada suhu ekstrem. Mengingat pembagian data uji dilakukan secara kronologis pada periode akhir, pola penyimpangan tersebut diduga terkait dengan indikasi anomali sensor di akhir tahun 2023 yang telah dijelaskan sebelumnya. Dengan demikian, kondisi data yang tidak ideal, khususnya terbatasnya variasi prediktor tekanan dan radiasi pada periode uji, menjadi salah satu faktor pembatas performa model.



Gambar 8. Perbandingan Temperatur Aktual dan Prediksi

Pada sub-periode awal data uji, 11 Januari sampai 2 Februari 2024, kurva prediksi yang ditunjukkan oleh garis merah putus-putus secara umum mengikuti pola kurva aktual yang digambarkan garis hitam dengan baik. Model juga mampu mereproduksi siklus harian temperatur secara konsisten sepanjang rentang waktu tersebut. Akan tetapi, pada tanggal 11–13 Januari model cenderung melakukan *overestimate* pada puncak harian, di mana prediksi mencapai sekitar 31°C sementara observasi hanya sekitar 28°C. Deviasi serupa juga terlihat pada beberapa puncak dan lembah harian lainnya. Meskipun demikian, model masih dapat menangkap pola fluktuasi temperatur harian secara keseluruhan. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang memadai dalam merepresentasikan dinamika deret waktu temperatur, namun akurasi prediksi pada nilai puncak harian masih perlu ditingkatkan.

Jika dibandingkan dengan studi sebelumnya, kinerja model dalam penelitian ini tergolong kompetitif dengan nilai RMSE 0,868°C, MAE 0,591°C, dan R^2 0,903, meskipun belum melampaui seluruh penelitian perbandingan. Nilai RMSE yang diperoleh lebih rendah daripada model ANN *backpropagation* milik Panji dan Sofwan sebesar 1,198°C. Namun nilai ini lebih tinggi dibanding model hibrida ARIMA-MLP dari Suciarti, dkk. dengan RMSE di bawah 0,20°C, serta model LSTM Supriyadi dengan RMSE 0,576°C. Perbedaan tersebut wajar mengingat penelitian ini menggunakan skema pembagian data kronologis atau *time*

series split yang lebih ketat dan realistis, sehingga mencegah kebocoran informasi dari masa depan ke data pelatihan, berbeda dengan skema pengujian pada beberapa studi perbandingan. Di samping itu, data yang digunakan mencakup periode lima tahun, 2020–2024, yang jauh lebih panjang dibanding studi sebelumnya yang umumnya hanya satu hingga dua tahun, serta memuat indikasi anomali sensor pada akhir 2023. Dengan kondisi pengujian yang lebih menantang tersebut, pencapaian R^2 sebesar 0,903 tetap menunjukkan performa yang andal. Hal ini juga menegaskan bahwa penambahan fitur *_temp_lag1_* pada MLP mampu mendekati kemampuan model deret waktu kompleks seperti LSTM, meskipun belum menyamai akurasi sepenuhnya, dengan efisiensi komputasi yang jauh lebih tinggi [15][16][17].

5. KESIMPULAN

Penelitian ini berhasil mengembangkan model *Multilayer Perceptron* (MLP) untuk mengimputasi nilai hilang pada data temperatur udara AWS Stasiun Meteorologi Kelas I Soekarno-Hatta tahun 2020–2024. Variabel yang digunakan sebagai prediktor meliputi kelembaban udara, tekanan udara, radiasi matahari, dan temperatur satu periode sebelumnya yaitu *_temp_lag1_*. Data dibagi menjadi data latih dan data uji dengan rasio 80:20 secara kronologis, serta dinormalisasi menggunakan *StandardScaler*. Model dengan dua lapisan tersembunyi masing-masing 150 dan 75 neuron menghasilkan kinerja yang baik

pada data uji, dengan RMSE 0,868°C, MAE 0,591°C, dan R^2 0,903. Capaian ini bersifat kompetitif jika dibandingkan dengan studi sebelumnya yang menggunakan arsitektur lebih kompleks seperti LSTM, namun tetap mempertahankan efisiensi komputasi khas MLP. Di sisi lain, ditemukan adanya indikasi anomali data pada akhir 2023 yang menjadi keterbatasan dan menyebabkan model cenderung *underestimate* pada suhu ekstrem. Untuk itu, disarankan penelitian berikutnya mengkaji penyebab anomali tersebut, menguji model pada skenario *missing value* beruntun dalam durasi panjang, serta membandingkannya dengan model berbasis deret waktu lain untuk mengevaluasi *trade-off* antara akurasi dan efisiensi komputasi secara lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] I. Irawati et al., "Pengenalan Instrumen Manual dan Digital untuk Mendukung Layanan Prakiraan Cuaca di Taman Alat BMKG Wilayah I DEO Sorong," *Abdimas: Papua J. Community Serv.*, vol. 7, no. 2, pp. 105–117, 2025, doi: 10.33506/pjcs.v7i2.4633.
- [2] S. A. Pramono and L. F. Muzayanah, "Kendali Mutu Data Suhu Udara Per Jam Menggunakan Metode Range Check di Stasiun Klimatologi Malang," *Jurnal Sains dan Teknologi Atmosfer*, vol. 2, no. 1, pp. 8–13, 2022.
- [3] A. Widiyanti and I. Pratama, "Penanganan Missing Values dan Prediksi Data Timbunan Sampah Berbasis Machine Learning," *RABIT: J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 2, pp. 242–251, 2024, doi: 10.36341/rabit.v9i2.4789.
- [4] D. Ramadhani, A. M. Soleh, and Erfiani, "Machine Learning-Based Univariate Time Series Imputation Method for Estimating Missing Values in Non-Stationary Data," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 307–320, 2024, doi: 10.20956/j.v21i1.36468.
- [5] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 52–62, 2023, doi: 10.37034/jidt.v5i2.324.
- [6] A. Wangwongchai, M. Waqas, P. Dechpichai, P. Thandar, and S. Ahmad, "Imputation of Missing Daily Rainfall Data: A Comparison Between Artificial Intelligence and Statistical Techniques," *MethodsX*, vol. 11, p. 102459, 2023, doi: 10.1016/j.mex.2023.102459.
- [7] C. Li, X. Ren, and G. Zhao, "Machine-Learning-Based Imputation Method for Filling Missing Values in Ground Meteorological Observation Data," *Algorithms*, vol. 16, no. 9, p. 422, 2023, doi: 10.3390/a16090422.
- [8] A. Ostovar, D. D. Davari, and M. Dzikuć, "Determinants of Design with Multilayer Perceptron Neural Networks: A Comparison with Logistic Regression," *Sustainability*, vol. 17, no. 6, p. 2611, 2025, doi: 10.3390/su17062611.
- [9] Y. Zhou, M. R. Bouadjeneq, and S. Aryal, "Developing Robust Methods to Handle Missing Data in Real-World Applications Effectively," *arXiv:2502.19635*, 2025.
- [10] A. Namdari and T. S. Durrani, "A Multilayer Feedforward Perceptron Model in Neural Networks for Predicting Stock Market Short-Term Trends," *Oper. Res. Forum*, pp. 1–30, 2021, doi: 10.1007/s43069-021-00071-2.
- [11] A. Barbu, "Training a Two-Layer ReLU Network Analytically," 2023.
- [12] I. Vajs, D. Drajić, N. Gligoric, I. Radovanovic, and I. Popovic, "Developing Relative Humidity and Temperature Corrections for Low-Cost Sensors Using Machine Learning," pp. 1–22, 2021.
- [13] I. Tri Handini, Y. Yulkifi, and Y. Darvina, "Rancang Bangun Sistem Pengukuran Tekanan Udara Menggunakan DT-Sense Barometric Pressure Berbasis Internet of Things dengan Display Smartphone," *J. Teor. dan Apl. Fis.*, vol. 8, no. 1, pp. 1–10, 2020, doi: 10.23960/jtaf.v8i1.2257.
- [14] Z. Feng, M. Leong, M. Amirul, J. Chuah, and F. Zhang, "Impacts of Solar Radiation Modification on Temperature Extremes and Heatwaves in Southeast Asia," *Weather Clim. Extrem.*, vol. 49, p. 100789, 2025, doi: 10.1016/j.wace.2025.100789.
- [15] S. Suciarti, M. R. T. Siregar, and A. Multi, "Estimasi Sensor Temperatur Udara Automatic Weather Station Menggunakan Algoritma Hybrid ARIMA-MLP," *Sainstech J. Penelit. dan Pengkaj. Sains dan Teknol.*, vol. 33, no. 3, pp. 42–48, 2023, doi: 10.37277/stch.v33i3.1740.
- [16] S. R. Panji and A. Sofwan, "Prediksi Sensor Cuaca Menggunakan ANN Backpropagation: Studi Kasus Suhu, Kelembaban, dan Tekanan Udara," *Sainstech J. Penelit. dan Pengkaj. Sains dan Teknol.*, vol. 35, no. 3, pp. 83–93, 2025, doi: 10.37277/stch.v35i3.2395.

- [17] E. Supriyadi, "Prediksi Parameter Cuaca Menggunakan Deep Learning Long-Short Term Memory (LSTM)," *J. Meteorol. dan Geofis.*, vol. 21, no. 2, pp. 55–67, 2020, doi: 10.31172/jmg.v21i2.619.
- [18] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.