ANALISIS PREDIKSI SUHU DENGAN HYBRID LSTM-GRU PADA DATASET BADAN METEOROLOGI STASIUN KEMAYORAN

Shofil Fuadi[1)], I Kadek Dwi Nuryana[2)]
[1)] Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: shofil.22102@mhs.unesa.ac.id
[2)] Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: dwinuryana@unesa.ac.id

Abstract
This study aims to analyze and compare the performance of LSTM, GRU, and Hybrid LSTM-GRU models in predicting ten-day mean air temperature at Kemayoran Meteorological Station and to develop an interactive dashboard for presenting the prediction results. The study follows the CRISP-DM framework and uses 1,096 daily weather observations from 20 November 2022 to 19 November 2025. The daily data are aggregated into 109 ten-day records using TAVG, RH_AVG, RR, SS, and FF_AVG. Model performance is evaluated through five-fold TimeSeriesSplit validation with three random seeds and testing on the final 20 targets. The results show that the mean validation MAE on the training data for LSTM, GRU, and Hybrid LSTM-GRU are 0.4855 °C, 0.4726 °C, and 0.5326 °C, respectively, while their testing MAE values are 0.5072 °C, 0.5117 °C, and 0.5291 °C. GRU is selected because it achieves the lowest mean validation MAE on the training data, although LSTM performs slightly better on the testing data. The selected GRU model is then implemented in a Streamlit dashboard consisting of four tabs and a one-step-ahead prediction feature using a 12-period window. These findings indicate that the additional complexity of the Hybrid LSTM-GRU model does not improve prediction accuracy on the dataset used.
Keywords: Temperature prediction; Ten-day period; LSTM; GRU; Hybrid LSTM-GRU.

Abstrak
Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja model LSTM, GRU, dan Hybrid LSTM-GRU dalam memprediksi suhu rata-rata udara dasarian di Stasiun Meteorologi Kemayoran serta mengembangkan dashboard interaktif untuk menyajikan hasil prediksi. Penelitian ini menggunakan kerangka CRISP-DM dengan memanfaatkan 1.096 data pengamatan cuaca harian dari Stasiun Meteorologi Kemayoran selama periode 20 November 2022 hingga 19 November 2025. Data harian tersebut diagregasi menjadi 109 observasi dasarian dengan melibatkan variabel TAVG, RH_AVG, RR, SS, dan FF_AVG. Kinerja model dievaluasi melalui validasi TimeSeriesSplit dengan lima fold dan tiga seed, kemudian diuji pada 20 target terakhir. Hasil penelitian menunjukkan bahwa rata-rata MAE validasi pada data training untuk model LSTM, GRU, dan Hybrid LSTM-GRU masing-masing sebesar 0,4855 °C, 0,4726 °C, dan 0,5326 °C, sedangkan MAE testing masing-masing sebesar 0,5072 °C, 0,5117 °C, dan 0,5291 °C. GRU dipilih karena menghasilkan rata-rata MAE validasi pada data training paling rendah, meskipun LSTM memperoleh MAE sedikit lebih rendah pada data testing. Model GRU terpilih kemudian diimplementasikan dalam dashboard Streamlit yang terdiri atas empat tab dan dilengkapi fitur prediksi satu langkah ke depan menggunakan window 12. Hasil tersebut menunjukkan bahwa penambahan kompleksitas pada model Hybrid LSTM-GRU belum meningkatkan akurasi prediksi pada dataset yang digunakan.
Kata Kunci: Prediksi suhu; Dasarian; LSTM; GRU; Hybrid LSTM-GRU.

PENDAHULUAN

Suhu udara menjadi salah satu indikator penting untuk memahami perubahan kondisi atmosfer, terutama di wilayah perkotaan yang mengalami perubahan penggunaan lahan dan aktivitas manusia secara intensif. Perubahan suhu tidak hanya terkait dengan kenyamanan, tetapi juga memengaruhi kebutuhan energi, pengelolaan air, kesehatan, dan perencanaan kegiatan masyarakat. Di Indonesia, BMKG melaporkan bahwa suhu rata-rata tahun 2024 mencapai 27,5 °C, meningkat dibandingkan rata-rata klimatologis 1991–2020 sebesar 26,7 °C (BMKG, 2025). Kondisi tersebut memperkuat kebutuhan terhadap metode prediksi yang dapat memanfaatkan data pengamatan meteorologi secara terukur.

Prediksi suhu merupakan persoalan deret waktu karena nilai pada suatu periode dipengaruhi oleh kondisi sebelumnya. Selain itu, perubahan suhu berkaitan dengan kelembapan, curah hujan, penyinaran matahari, dan angin. Hubungan antarvariabel tersebut dapat bersifat tidak linier, sehingga penggunaan model deep learning recurrent menjadi salah satu pendekatan yang relevan. Long Short-Term Memory (LSTM) memanfaatkan mekanisme gerbang untuk mempertahankan informasi temporal (Hochreiter & Schmidhuber, 1997), sedangkan Gated Recurrent Unit (GRU) menawarkan struktur lebih ringkas dengan mekanisme reset dan update (Cho et al., 2014). Keduanya juga dapat digabungkan dalam arsitektur Hybrid LSTM-GRU.

Penelitian sebelumnya menunjukkan hasil yang tidak selalu seragam. Mustofa dan Setiawan (2024) menemukan bahwa LSTM dan GRU mampu memprediksi indikator iklim di Indonesia dengan tingkat kesalahan yang berdekatan. Di sisi lain, Pangaribuan et al. (2024) melaporkan kinerja Hybrid LSTM-GRU yang baik pada data meteorologi beresolusi sepuluh menit. Perbedaan karakteristik data dan horizon prediksi mengisyaratkan bahwa kinerja suatu arsitektur perlu diuji pada konteks pengamatan tertentu, bukan ditentukan semata-mata oleh kompleksitas jaringannya. Furizal et al. (2024) juga menekankan peran karakteristik dataset dan konfigurasi pelatihan dalam perbandingan kedua arsitektur recurrent.

Penggunaan data meteorologi Kemayoran juga memiliki pijakan pada studi Supriyadi (2021) yang menerapkan LSTM untuk parameter cuaca di lokasi tersebut, tetapi hanya mencakup dua bulan pengamatan. Gunawan et al. (2020) menunjukkan bahwa LSTM dapat digunakan untuk prediksi suhu di Indonesia, sementara Yunita et al. (2025) membahas perbandingan beragam arsitektur recurrent dan Hybrid pada deret waktu. Namun, perbandingan LSTM, GRU, dan Hybrid secara bersamaan pada rentang tiga tahun dengan agregasi dasarian dan keluaran dashboard masih menjadi fokus khusus yang ingin dilengkapi melalui penelitian ini.

Kajian ini berfokus pada Stasiun Meteorologi Kemayoran, Jakarta, menggunakan periode pengamatan tiga tahun dan satuan dasarian atau sekitar sepuluh harian. Perbandingan dilakukan terhadap LSTM, GRU, dan Hybrid LSTM-GRU dengan data dan prosedur evaluasi yang sama. Penelitian bertujuan (1) membandingkan tingkat kesalahan prediksi ketiga model; (2) menelaah hasil validasi dan pengujian akhir secara terpisah; serta (3) menyajikan model terpilih dan hasil eksperimen melalui dashboard interaktif berbasis Streamlit. Penggunaan dasarian dipilih agar pembacaan data selaras dengan interval analisis yang lazim digunakan dalam klimatologi (World Meteorological Organization, 2011).

METODE PENELITIAN

1. Rancangan penelitian.

Alur kerja mengikuti enam tahap CRISP-DM, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment (Chapman et al., 2000). Penentuan model terbaik dilakukan berdasarkan data training agar data testing tetap berfungsi sebagai evaluasi akhir yang terpisah. Seluruh model memprediksi suhu rata-rata pada satu dasarian berikutnya (one-step-ahead).

2. Sumber dan variabel data.

Dataset berasal dari Stasiun Meteorologi Kemayoran (WMO 96745) untuk 20 November 2022 sampai 19 November 2025. Data memuat 1.096 observasi harian. Lima variabel yang digunakan adalah suhu rata-rata (TAVG) sebagai fitur historis sekaligus target, kelembapan rata-rata (RH_AVG), curah hujan (RR), lama penyinaran matahari (SS), dan kecepatan angin rata-rata (FF_AVG). Penggunaan beberapa variabel memungkinkan model menangkap informasi temporal sekaligus keterkaitan antarunsur iklim.

3. Penyiapan dan agregasi data.

Tanggal diurutkan secara kronologis. Nilai “-”, 8888, dan 9999 diperlakukan sebagai data hilang, sedangkan nilai nol dipertahankan sebagai pengamatan valid. RR memiliki data hilang terbanyak, yaitu 217 pengamatan atau 19,80%. Data kemudian diagregasi menjadi 109 observasi dasarian: TAVG, RH_AVG, SS, dan FF_AVG dihitung sebagai rata-rata, sedangkan RR dijumlahkan hanya dari hari dengan data valid. Karena sebagian pengamatan hujan hilang, jumlah RR dasarian tidak selalu mewakili total hujan lengkap. Fitur kosong setelah agregasi ditangani secara kronologis dengan forward fill dan median data training sebagai cadangan. Nilai ekstrem meteorologis tidak otomatis dihapus.

 ISSN 3063-864X (E)

Pemeriksaan data menunjukkan bahwa kualitas tiap atribut tidak sama. Data suhu rata-rata memiliki empat pengamatan hilang (0,36%), kelembapan dua (0,18%), lama penyinaran delapan (0,73%), dan kecepatan angin tidak memiliki nilai hilang. Sebaliknya, curah hujan mengandung 217 nilai hilang, yang berasal dari tanda "-" dan kode khusus. Karena angka nol menyatakan hari tanpa hujan, menyamakan nilai hilang dengan nol akan berisiko mengecilkan besaran hujan dan mengubah pola hubungan dengan suhu. Oleh sebab itu, jumlah hari valid tetap dicatat ketika data curah hujan diringkas ke dasarian. Kebijakan ini penting agar nilai yang tersaji dalam hasil pengolahan dapat ditafsirkan secara tepat.

4. Pembentukan sequence dan pembagian data.

Panjang sliding window yang diuji adalah 3, 6, 9, dan 12 dasarian. Agar periode target dapat dibandingkan secara adil, semua skenario diselaraskan pada 97 target yang memenuhi kebutuhan window terpanjang.

Dari jumlah tersebut, 77 target digunakan untuk training dan 20 target terakhir untuk testing pada 1 Mei–19 November 2025. Data training dinilai menggunakan TimeSeriesSplit lima fold dengan expanding window dan tiga seed untuk setiap model final. Scaler fitur dan target di-fit hanya pada bagian training setiap fold, lalu hasil prediksi dikembalikan ke satuan °C sebelum menghitung kesalahan.

5. Arsitektur dan evaluasi.

Model LSTM tunggal memakai satu layer LSTM, dropout, dan Dense(1). Model GRU tunggal memakai satu layer GRU, dropout, dan Dense(1).

Model Hybrid menyusun layer LSTM yang mengeluarkan sekuens, diikuti GRU dan Dense(1). Pencarian hyperparameter meliputi jumlah unit, dropout, learning rate, batch size, optimizer, dan patience. Pelatihan meminimalkan MSE dengan early stopping; pemilihan konfigurasi mengutamakan rata-rata Mean Absolute Error (MAE) pada validasi. Mean Squared Error (MSE) dan Root Mean Squared Error (RMSE) menjadi metrik pendukung. Secara umum, MAE adalah rata-rata $|y-\hat{y}|$, MSE adalah rata-rata $(y-\hat{y})^2$, dan RMSE merupakan akar MSE. Semakin kecil nilainya, semakin kecil kesalahan prediksi.

Prosedur pemisahan temporal perlu diperhatikan karena data meteorologi mempunyai urutan waktu yang tidak dapat dipertukarkan seperti pada data tabular acak. Jika scaler atau proses imputasi menggunakan informasi dari periode testing ketika model masih dilatih, hasil evaluasi berpotensi menjadi terlalu optimistis. Dalam eksperimen final ini, parameter normalisasi ditentukan dari training pada setiap fold dan diterapkan kembali pada periode validasi yang menyusulnya. Early stopping memakai bagian akhir training di dalam fold, kemudian model dilatih ulang sesuai jumlah epoch terpilih sebelum menilai validasi luar. Proses tersebut menjaga agar pencarian konfigurasi tidak bergantung pada target di masa depan.

6. Pengujian statistik dan implementasi.

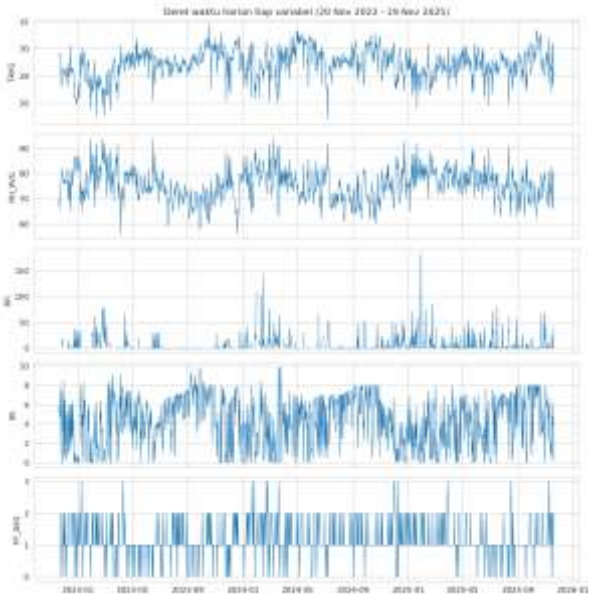
Perbedaan antar model ditinjau melalui uji Friedman eksploratif pada rata-rata MAE tiga seed per fold. Karena fold TimeSeriesSplit memiliki bagian training yang saling tumpang tindih, hasil uji diperlakukan sebagai informasi tambahan, bukan bukti independen yang konklusif. Model yang telah dipilih dari validasi selanjutnya diintegrasikan dalam dashboard Streamlit yang menampilkan data historis, prediksi, serta perbandingan hasil evaluasi.

HASIL DAN PEMBAHASAN

1. Karakteristik Data

Data harian menunjukkan suhu rata-rata (TAVG) sebesar 28,89 °C dengan standar deviasi 1,07 °C dan rentang 24,80–31,80 °C. Rata-rata kelembapan mencapai 75,58%, lama penyinaran matahari 4,27 jam, dan kecepatan angin 1,15 m/s. Curah hujan jauh lebih berfluktuasi, dengan rata-rata 7,34 mm dan standar deviasi 16,75 mm. Pola tersebut penting untuk dipahami karena model mempelajari perubahan suhu dari kombinasi riwayat variabel, bukan dari nilai suhu saja.

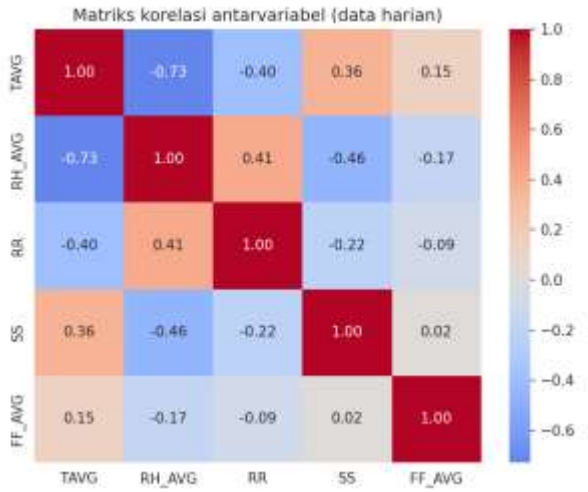
Grafik deret waktu harian pada Gambar 1 menunjukkan bahwa suhu bergerak pada rentang yang relatif sempit dibandingkan curah hujan yang memiliki sejumlah lonjakan. Perubahan kelembapan dan penyinaran juga terlihat sepanjang periode pengamatan. Grafik asli ini membantu menjelaskan mengapa data harian diagregasi ke dasarian: arah perubahan dapat dibaca lebih stabil tanpa menghilangkan riwayat antarvariabel yang digunakan dalam pemodelan.



Gambar 1. Deret waktu harian lima variabel meteorologi.
Sumber : Hasil Pengolahan Data Peneliti (2026).

 ISSN 3063-864X (E)

Matriks korelasi pada Gambar 2 memperlihatkan hubungan TAVG dengan RH_AVG sebesar −0,73, RR sebesar −0,40, SS sebesar 0,36, dan FF_AVG sebesar 0,15. Kelembapan dan curah hujan cenderung berkaitan dengan suhu yang lebih rendah pada pengamatan ini, sedangkan penyinaran matahari berkorelasi positif. Namun, korelasi hanya menunjukkan asosiasi; hasilnya tidak cukup untuk menyatakan hubungan sebab-akibat.

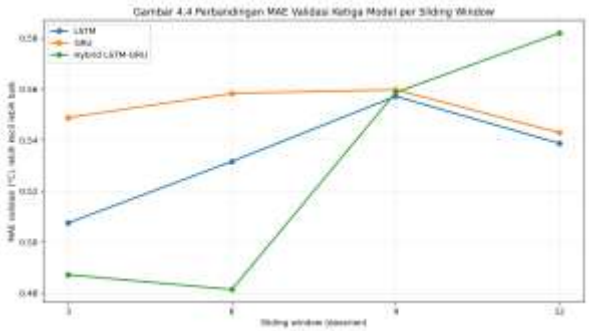


Gambar 2. Matriks korelasi variabel meteorologi.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Bentuk kurva pada Gambar 1 memperlihatkan sejumlah perubahan suhu dalam jangka pendek, termasuk penurunan mendadak pada beberapa tanggal. Pada data dasarian, fluktuasi singkat semacam itu menjadi lebih halus setelah dirata-ratakan. Keuntungan agregasi adalah pola antardasarian lebih mudah dipelajari, tetapi informasi ekstrem harian tidak lagi terlihat secara utuh. Karena itu, hasil model dalam penelitian ini perlu dimaknai sebagai prediksi suhu rata-rata dasarian, bukan sebagai prakiraan suhu maksimum, minimum, ataupun kejadian cuaca ekstrem harian.

2. Perbandingan Sliding Window

Empat ukuran window diuji pada tahap awal untuk menemukan rentang riwayat yang sesuai. Gambar 3 menyajikan MAE validasi pada tahap ini. Hybrid LSTM-GRU mencapai nilai terendah pada window 6 sebesar 0,4814 °C, sedangkan hasil terbaik tahap awal LSTM berada pada window 3 (0,5075 °C) dan GRU pada window 12 (0,5429 °C). Grafik tahap awal ini tidak dijadikan satu-satunya dasar penentuan model akhir karena tuning dilanjutkan dengan perubahan unit dan hyperparameter lainnya.



Gambar 3. Perbandingan MAE validasi tahap awal menurut sliding window.
Sumber : Hasil Pengolahan Data Peneliti (2026).

3. Konfigurasi Terbaik

Setelah tuning, LSTM menggunakan window 3 dan 64 unit, GRU menggunakan window 12 dan 16 unit, sedangkan Hybrid LSTM-GRU menggunakan window 6 dengan 64 unit LSTM serta 16 unit GRU. GRU menjadi model dengan parameter paling sedikit, yaitu 1.121, dibandingkan LSTM 17.985 dan Hybrid 21.873 parameter (Tabel 1). Perbandingan jumlah parameter menunjukkan perbedaan kompleksitas arsitektur, tetapi tidak dengan sendirinya menentukan akurasi.

Tabel 1. Konfigurasi final ketiga model.

Model	Window	Unit recurrent	Parameter
LSTM	3	64	17.985
GRU	12	16	1.121
Hybrid LSTM-GRU	6	64 LSTM + 16 GRU	21.873

4. Hasil Validasi Data Training

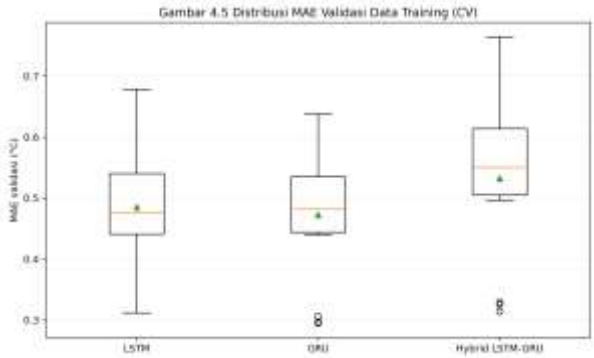
Evaluasi lima fold dengan tiga seed menghasilkan 15 nilai pengukuran per model. Tabel 2 menunjukkan bahwa GRU memperoleh rata-rata MAE terendah sebesar 0,4726 °C, diikuti LSTM 0,4855 °C dan Hybrid 0,5326 °C. GRU juga memiliki RMSE validasi terendah sebesar 0,5978 °C. Walaupun LSTM sedikit lebih stabil dilihat dari standar deviasi MAE (0,1027 dibandingkan GRU 0,1064), perbedaan rata-rata kesalahan menjadi dasar pemilihan GRU sesuai kriteria yang ditentukan sejak awal.

Tabel 2. Hasil validasi training (5 fold × 3 seed).

Model	MAE (°C)	SD MAE	RMSE (°C)
LSTM	0,4855	0,1027	0,6029
GRU	0,4726	0,1064	0,5978
Hybrid LSTM-GRU	0,5326	0,1266	0,6549

 ISSN 3063-864X (E)

Sebaran nilai validasi (Gambar 4) memperlihatkan variasi hasil antar-fold dan seed. Karena masing-masing percobaan berangkat dari periode dan inisialisasi bobot yang berbeda, nilai rata-rata saja belum cukup menggambarkan kestabilan model. Pada data ini, arsitektur Hybrid yang lebih besar belum menunjukkan keuntungan pada metrik MAE maupun RMSE.



Gambar 4. Distribusi MAE validasi pada data training.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Perbedaan kinerja validasi juga perlu dilihat bersama kompleksitas dan proses pelatihan. Pada hasil eksekusi yang terdokumentasi dalam skripsi, rata-rata waktu pelatihan per proses tercatat sekitar 11,54 detik untuk LSTM, 7,45 detik untuk GRU, dan 6,23 detik untuk Hybrid. Lama pelatihan dipengaruhi konfigurasi dan kondisi eksekusi, sehingga angka tersebut tidak cukup untuk menyimpulkan bahwa satu arsitektur selalu lebih cepat. Namun, jumlah parameter yang jauh lebih sedikit pada GRU tetap menjadi alasan praktis untuk mempertimbangkan kesederhanaan implementasinya, selama tingkat kesalahan prediksinya dapat diterima.

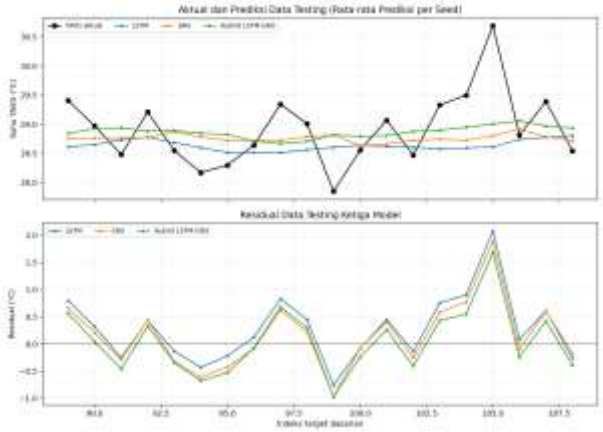
5. Evaluasi pada Data Testing

Setelah model dipilih, seluruh arsitektur diuji pada 20 target dengan tanggal yang sama, yaitu 1 Mei–19 November 2025. Hasilnya memperlihatkan pola yang sedikit berbeda dari validasi. LSTM memperoleh MAE terendah sebesar 0,5072 °C, diikuti GRU 0,5117 °C dan Hybrid 0,5291 °C (Tabel 3). Selisih MAE LSTM dan GRU hanya 0,0045 °C. Namun, dari sisi RMSE testing, GRU memiliki nilai paling rendah sebesar 0,6611 °C. Keputusan menggunakan GRU tidak diubah setelah melihat testing, karena data tersebut sejak awal disisihkan untuk mengukur generalisasi, bukan untuk memilih konfigurasi.

Tabel 3. Evaluasi pada 20 target data testing.

Model	MAE (°C)	MSE (°C²)	RMSE (°C)
LSTM	0,5072	0,4607	0,6784
GRU	0,5117	0,4394	0,6611
Hybrid LSTM-GRU	0,5291	0,4643	0,6741

Grafik pada Gambar 5 membandingkan suhu aktual, prediksi ketiga model, dan residualnya. Model dapat mengikuti tingkat suhu rata-rata secara umum, tetapi masih cenderung meredam lonjakan dan penurunan yang tajam. Grafik menggunakan rata-rata prediksi dari tiga seed, sedangkan metrik pada Tabel 3 merupakan rata-rata metrik yang dihitung per seed. Oleh karena itu, nilai kesalahan yang dihitung ulang dari garis rata-rata prediksi pada gambar tidak harus identik dengan tabel.



Gambar 5. Suhu aktual, prediksi, dan residual pada data testing.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Terdapat dua hal yang perlu diperhatikan saat membaca Tabel 2 dan Tabel 3. Pertama, konfigurasi yang optimal pada data training belum tentu memberikan urutan kesalahan yang sama pada data testing, terutama ketika sampel testing hanya mencakup 20 dasarian. Kedua, perbandingan model dilakukan dengan membedakan metrik primer MAE dari metrik pendukung MSE dan RMSE. Pada testing, misalnya, GRU memiliki RMSE lebih rendah daripada LSTM meskipun MAE-nya sedikit lebih tinggi. Dengan demikian, penentuan model didasarkan pada kriteria penelitian yang konsisten, sementara metrik lain tetap dilaporkan secara transparan agar pembaca dapat menilai karakter kesalahannya.

6. Uji Statistik dan Pembahasan

Uji Friedman eksploratif terhadap rata-rata MAE tiga seed pada lima fold memperoleh statistik 2,8000 dengan p-value 0,2466. Pada taraf 0,05, hasil ini belum memberikan bukti yang cukup bahwa perbedaan kinerja ketiga model signifikan secara statistik. Nilai tersebut juga tidak membuktikan bahwa ketiganya setara. Lebih jauh, unit analisis antar-fold tidak sepenuhnya independen karena rentang training dalam TimeSeriesSplit saling bertumpang tindih. Karena alasan itu, uji diposisikan sebagai penjelas tambahan, bukan sebagai dasar utama pemilihan model.

 ISSN 3063-864X (E)

Secara empiris, GRU dengan 16 unit dan window 12 memberikan titik keseimbangan yang baik antara tingkat kesalahan dan kesederhanaan jaringan pada dataset ini. Model Hybrid menggabungkan dua jenis layer recurrent dan memiliki parameter hampir dua puluh kali lebih banyak daripada GRU, tetapi justru mencatat MAE validasi yang lebih tinggi. Temuan tersebut berbeda dari hasil Pangaribuan et al. (2024) pada data meteorologi beresolusi sepuluh menit serta Li et al. (2026) pada mikroklimat rumah kaca, yang menunjukkan kinerja baik arsitektur Hybrid. Perbedaan ini masuk akal mengingat resolusi waktu, jumlah observasi, dan lingkungan pengamatan tidak sama. Temuan juga sejalan dengan kajian Furizal et al. (2024) bahwa tidak terdapat satu arsitektur recurrent yang otomatis unggul di semua kondisi.

Keterbatasan utama penelitian terletak pada jumlah data dasarian yang relatif kecil, yakni 109 observasi, serta kelengkapan curah hujan yang belum merata. Selain itu, hasil pada satu stasiun belum tentu mewakili wilayah lain. Temuan ini sebaiknya dipahami sebagai hasil komparasi empiris untuk periode, fitur, dan prosedur evaluasi yang digunakan, bukan sebagai klaim keunggulan universal GRU maupun ketidakmampuan arsitektur Hybrid.

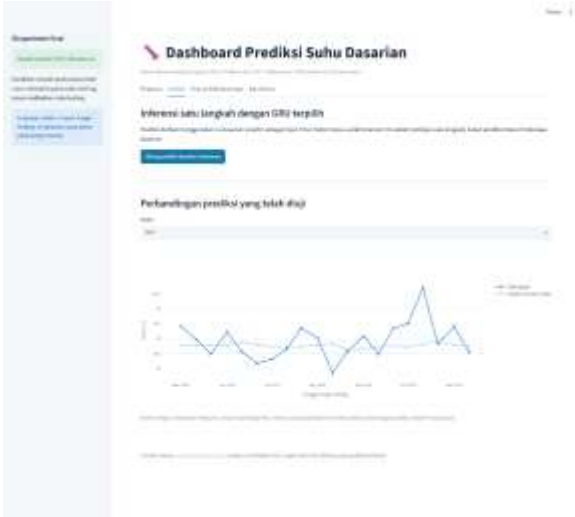
7. Implementasi Dashboard

Model terpilih diintegrasikan ke dashboard berbasis Streamlit untuk mempermudah pembacaan hasil penelitian. Dashboard memiliki empat tab: Ringkasan, Prediksi, Evaluasi dan Perbandingan, serta Data Historis. Tab Ringkasan menampilkan identitas GRU sebagai model terpilih, nilai MAE, dan tren suhu. Tab Prediksi dirancang untuk inferensi satu dasarian ke depan dari 12 dasarian terakhir dan menampilkan perbandingan dengan hasil testing. Tab evaluasi merangkum metrik ketiga model, sedangkan tab historis menyajikan deret data serta indikator kelengkapannya. Contoh tampilan tab Ringkasan dari skripsi ditunjukkan pada Gambar 6.



Gambar 6. Tampilan tab Ringkasan pada dashboard Streamlit.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Tab Prediksi pada Gambar 7 memperlihatkan rancangan inferensi satu langkah dan grafik evaluasi terhadap data testing. Informasi ini penting karena pengguna dapat membedakan ramalan satu periode mendatang dari hasil pengujian model pada periode yang sudah memiliki pengamatan aktual. Dengan menempatkan hasil evaluasi bersama visualisasi, dashboard berfungsi sebagai media penjelasan, bukan sekadar alat yang menghasilkan satu angka prakiraan.



Gambar 7. Tampilan tab Prediksi dashboard Streamlit.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Tab Evaluasi dan Perbandingan pada Gambar 8 menyajikan rekap hasil validasi silang pada data training, distribusi kesalahan prediksi ketiga model, serta perbandingan MAE, konfigurasi, dan residual pada data testing. Informasi ini membantu pembaca menelaah alasan pemilihan GRU berdasarkan hasil validasi sekaligus membandingkannya secara terpisah dengan hasil pengujian akhir.



Gambar 8. Tab Evaluasi dan Perbandingan Streamlit.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Tab Data Historis pada Gambar 9 melengkapi visualisasi dengan tabel observasi dasarian dan indikator kelengkapan data. Penyajian ini berguna untuk menjelaskan asal masukan model sekaligus memberi konteks ketika nilai suatu dasarian perlu dibaca dengan hati-hati, terutama untuk curah hujan yang berasal dari pengamatan harian tidak lengkap. Tampilan histori dan grafik hasil evaluasi disediakan dalam satu aplikasi, sehingga pembaca dapat menghubungkan data, konfigurasi model, dan hasil prediksi tanpa harus membuka file eksperimen secara terpisah.



Gambar 9. Tampilan tab Data Historis dashboard.
Sumber : Hasil Pengolahan Data Peneliti (2026).

Artefak GRU dan scaler telah melalui verifikasi inferensi lokal, sementara keempat tab dashboard telah terdokumentasi tampil. Meskipun demikian, bukti pengujian keluaran setelah tombol prediksi ditekan belum tercantum dalam skripsi. Oleh karena itu, pengujian interaksi tersebut tetap diperlukan sebelum dashboard digunakan lebih luas. Pemisahan informasi validasi dan testing dalam dashboard juga membantu pengguna memahami alasan GRU dipilih meskipun MAE testing LSTM sedikit lebih rendah.

KESIMPULAN

Hasil penelitian menunjukkan bahwa LSTM, GRU, dan Hybrid LSTM-GRU dapat digunakan untuk memprediksi suhu rata-rata dasarian berdasarkan data multivariat Stasiun Meteorologi Kemayoran. Pada validasi training lima fold dan tiga seed, GRU memperoleh MAE paling rendah sebesar 0,4726 °C, diikuti LSTM 0,4855 °C dan Hybrid 0,5326 °C. GRU dengan window 12 dan 16 unit dipilih sebagai model akhir sebelum hasil testing diperiksa. Pada testing, LSTM mencatat MAE terendah sebesar 0,5072 °C, sedikit di bawah GRU 0,5117 °C. Uji Friedman eksploratif dengan $p = 0{,}2466$ belum menunjukkan bukti perbedaan yang signifikan. Oleh karena itu, hasil perbandingan ini menegaskan perlunya pemilihan model melalui prosedur validasi yang konsisten dan pembacaan hasil testing secara terpisah.

Model GRU terpilih telah diterapkan dalam dashboard Streamlit untuk menampilkan tren suhu, prediksi, evaluasi model, dan data historis. Pengembangan selanjutnya disarankan mencakup perluasan periode serta wilayah pengamatan, pengujian hyperparameter yang lebih luas, penambahan variabel iklim yang relevan, dan dokumentasi uji interaktif dashboard, terutama setelah tombol prediksi dijalankan.

DAFTAR PUSTAKA

BMKG. (2025). Anomali Suhu Udara Rata-Rata Tahun 2024.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Cho, K., Van Merriënboer, B., Gulçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. Proceedings of EMNLP, 1724–1734.

Furizal, F., Fawait, A. B., Maghfiroh, H., Ma'arif, A., Firdaus, A. A., & Suwarno, I. (2024). Long Short-Term Memory vs Gated Recurrent Unit: A literature review on the performance of deep learning methods in temperature time series forecasting. International Journal of Robotics and Control Systems, 4(3), 1506–1526. https://doi.org/10.31763/ijrcs.v4i3.1546

Gunawan, P. H., Munandar, D., & Farabiba, A. Z. (2020). Long short-term memory approach for predicting air temperature in Indonesia. Jurnal Online Informatika, 161–168. https://doi.org/10.15575/join.v5i2.551

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Li, B., Wang, Z., Guo, Y., Zhou, X., Qiao, B., & Han, L. (2026). LSTM-GRU hybrid model for multi-layer microclimate prediction in solar greenhouse. Scientific Reports, 16(1), 18625. https://doi.org/10.1038/s41598-026-49691-w

Mustofa, A., & Setiawan, H. (2024). Perbandingan metode deep learning dengan model LSTM dan GRU untuk prediksi perubahan iklim. Journal of Computer System and Informatics (JoSYC), 5(4), 868–882. https://doi.org/10.47065/josyc.v5i4.5671

Pangaribuan, A. R., Yuniarno, E. M., & Rachmadi, R. F. (2024). Air temperature prediction with LSTM and LSTM-GRU: A two-year meteorological data. 2024 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), 1–5. https://doi.org/10.1109/CENIM64038.2024.10882637

Supriyadi, E. (2021). Prediksi parameter cuaca menggunakan deep learning long-short term memory (LSTM).

World Meteorological Organization. (2011). Guide to climatological practices (WMO-No. 100). World Meteorological Organization.

Yunita, A., Pratama, M. I., Almuzakki, M. Z., Ramadhan, H., Akhir, E. A. P., Firdausiah Mansur, A. B., & Basori, A. H. (2025). Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. MethodsX, 15, 103462. https://doi.org/10.1016/j.mex.2025.103462