

PERAMALAN PENYINARAN MATAHARI PER JAM SATU HARI KE DEPAN MENGGUNAKAN MODEL LONG SHORT-TERM MEMORY (LSTM)

Dwinugroho Putro Utomo^{1*}

¹Departemen Pendidikan Teknik Elektro, Fakultas Teknik, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; Telp. (0274) 586168

Keywords:

peramalan iradiasi matahari;
LSTM;
K-means;
Bayesian Optimization;
one-day ahead.

Correspondent Email:

dwinugrohoptutomo@uny.ac.id

Abstrak. Energi surya merupakan sumber energi terbarukan yang potensial, namun variabilitas penyinaran matahari akibat kondisi cuaca menimbulkan tantangan dalam integrasi ke sistem tenaga listrik. Oleh karena itu, peramalan iradiasi matahari dengan akurasi tinggi menjadi penting untuk mendukung stabilitas dan efisiensi operasi jaringan. Penelitian ini mengusulkan model Long Short-Term Memory (LSTM) untuk peramalan iradiasi per jam satu hari ke depan dengan hanya menggunakan data historis Global Horizontal Irradiance (GHI) sebagai input. Proses penelitian meliputi pra-pemrosesan data, rekayasa fitur melalui klusterisasi K-Means untuk identifikasi pola cuaca, pemanfaatan data historis beberapa hari sebelumnya, serta transformasi waktu siklikal. Optimasi hyperparameter dilakukan menggunakan Bayesian Optimization untuk memperoleh konfigurasi model terbaik. Hasil evaluasi menunjukkan model mencapai R^2 sebesar 0,7937, RMSE 87,71 W/m², MAE 59,34 W/m², dan MAPE 28,75%. Analisis visual juga memperlihatkan kemampuan model mengikuti pola siklus harian dan berkorelasi kuat dengan data aktual. Temuan ini menegaskan bahwa data GHI historis saja sudah cukup menghasilkan prediksi yang reliabel, khususnya di wilayah dengan keterbatasan infrastruktur, serta dapat menjadi dasar pengembangan metode prediksi yang lebih efisien.

Abstract. Solar energy is a promising renewable source. However, the variability of solar irradiance due to weather conditions poses challenges for integration into power systems. Accurate irradiance forecasting is therefore essential to support grid stability and operational efficiency. This study proposes a Long Short-Term Memory (LSTM) model for one-day-ahead hourly solar irradiance forecasting. The model uses only historical Global Horizontal Irradiance (GHI) data as input. The research workflow includes data preprocessing and feature engineering through K-Means clustering to identify weather patterns. It also uses historical data from previous days and employs cyclic time transformation. Hyperparameter optimization was performed using Bayesian Optimization to obtain the best model configuration. Evaluation results show an R^2 of 0.7937, RMSE of 87.71 W/m², MAE of 59.34 W/m², and MAPE of 28.75%. Visual analysis demonstrates the model's ability to follow daily cyclic patterns and shows strong correlation with actual data. The findings highlight that historical GHI data alone are sufficient to produce reliable forecasts, particularly in regions with limited infrastructure. This can serve as a basis for developing more efficient forecasting approaches.



Copyright © [JITET](http://jitet.uny.ac.id) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

1. PENDAHULUAN

Energi matahari kini dipandang sebagai salah satu solusi utama dalam menghadapi tantangan penyediaan energi global yang berkelanjutan, sekaligus menekan dampak negatif terhadap lingkungan [1]. Tren dunia dalam kebijakan energi yang mengedepankan keberlanjutan telah memperkuat posisi teknologi surya, khususnya fotovoltaik (PV), sebagai komponen vital dalam pengembangan sistem energi bersih [2]. Meskipun potensinya besar, pemanfaatan energi surya masih dihadapkan pada berbagai kendala, terutama karena sifatnya yang sangat bergantung pada kondisi cuaca yang tidak dapat diprediksi secara konsisten. Hal ini menyebabkan fluktuasi produksi daya yang menyulitkan integrasi ke dalam sistem jaringan listrik modern [3], [4]. Ketidakpastian ini telah mendorong kebutuhan mendesak akan peningkatan akurasi dalam memprediksi penyinaran matahari global horizontal (Global Horizontal Irradiance, GHI) [5].

Sifat penyinaran matahari yang tidak menentu membawa tantangan besar dalam operasionalisasi jaringan, khususnya terkait stabilitas sistem dan pengelolaan energi [6]. Karena penyinaran matahari merupakan faktor utama dalam pembangkitan energi surya, pengaruh besar dari kondisi atmosfer menjadikannya sangat penting untuk diprediksi, baik dalam konteks perencanaan maupun pemeliharaan sistem [7]. Variabilitas dalam pembangkitan daya PV dapat mengganggu keandalan sistem kelistrikan secara keseluruhan, berdampak pada stabilitas, perencanaan operasional, serta efisiensi ekonomi dalam distribusi energi [8], [9]. Ketidakpastian ini juga menyulitkan operator jaringan dalam menjaga keseimbangan antara pasokan dan permintaan secara real-time [10].

Oleh karena itu, peramalan penyinaran matahari yang akurat sangat diperlukan untuk mendukung integrasi energi surya ke dalam jaringan listrik, khususnya dalam cakupan waktu jangka pendek seperti prediksi per jam untuk satu hari ke depan (one-day-ahead hourly forecasting) [11]. Prediksi semacam ini memungkinkan pengelolaan energi yang lebih efisien, penjadwalan sistem penyimpanan, serta optimalisasi alokasi sumber daya bagi operator sistem maupun pelaku perdagangan energi [12]. Beragam pendekatan telah dikembangkan untuk menjawab kebutuhan tersebut, banyak di antaranya mengandalkan beberapa data masukan seperti informasi cuaca, citra satelit, atau hasil prakiraan meteorologi numerik [11], [13]. Namun demikian, penggunaan berbagai input tersebut tidak terlepas

dari tantangan, seperti masalah kualitas data, kompleksitas model, dan tingginya variabilitas lingkungan [14], [15].

Sebagai alternatif dari pendekatan multi-input, studi ini mengusulkan metode peramalan berbasis input tunggal, yaitu hanya menggunakan data historis GHI yang dikumpulkan dari sensor permukaan. Dengan membatasi input hanya pada nilai-nilai GHI aktual, penelitian ini bertujuan untuk menguji sejauh mana kekuatan prediktif dari deret waktu penyinaran itu sendiri—tanpa melibatkan fitur tambahan dari luar. Strategi ini diperuntukkan untuk dapat diterapkan di lokasi-lokasi yang hanya memiliki infrastruktur sensor sederhana.

Untuk menangkap pola temporal dan hubungan nonlinier dalam data GHI, kami menggunakan arsitektur Long Short-Term Memory (LSTM), yaitu salah satu model deep learning yang dirancang untuk tugas-tugas prediksi berurutan [16]. LSTM unggul dalam mengolah data historis dan mengenali variasi temporal yang kompleks, sehingga sangat sesuai untuk aplikasi peramalan penyinaran matahari [17]. Temuan utama dalam penelitian ini menunjukkan bahwa model LSTM memberikan hasil yang lebih baik dibandingkan arsitektur lain, dengan capaian nilai R-squared tertinggi dan Root Mean Square Error (RMSE) terendah [18]. Sejumlah studi juga menegaskan bahwa model deep learning seperti LSTM sangat efektif dalam memprediksi output PV jangka pendek [19]. Kemampuannya dalam mempertahankan dependensi jangka panjang tanpa mengalami masalah vanishing gradient menjadikannya lebih unggul dibandingkan metode peramalan tradisional [16], [20].

Dengan mempertimbangkan tantangan yang melekat pada variabilitas penyinaran matahari dan keterbatasan pendekatan berbasis multi-input, penelitian ini diarahkan untuk mengevaluasi efektivitas model LSTM dalam melakukan peramalan penyinaran matahari harian per jam dengan hanya menggunakan data historis GHI. Pendekatan ini tidak hanya menyederhanakan sistem input, tetapi juga berpotensi meningkatkan ketersediaan model prediktif di wilayah dengan infrastruktur data yang terbatas. Melalui rancangan dan pengujian model LSTM yang diusulkan, diharapkan kontribusi nyata dapat diberikan bagi integrasi energi surya yang lebih andal dan efisien dalam sistem kelistrikan modern.

2. TINJAUAN PUSTAKA

Peramalan iradiasi matahari menjadi aspek penting dalam integrasi energi surya, karena

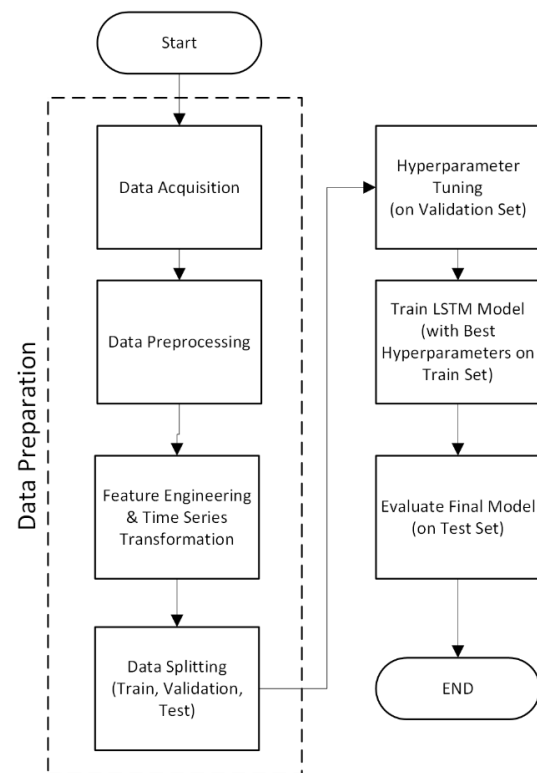
variabilitas cuaca dapat menyebabkan fluktuasi daya keluaran sistem fotovoltaik (PV) dan memengaruhi stabilitas jaringan listrik [5], [7]. Salah satu parameter yang umum digunakan adalah *Global Horizontal Irradiance (GHI)*, yang merepresentasikan intensitas penyinaran matahari pada bidang horizontal [11].

Metode berbasis *time series forecasting* telah banyak diterapkan untuk memprediksi iradiasi matahari. Namun, model tradisional seperti regresi linear dan *autoregressive models* seringkali terbatas dalam menangkap pola nonlinier yang kompleks [14], [15]. Untuk itu, model berbasis *deep learning* seperti *Long Short-Term Memory (LSTM)* menjadi alternatif yang lebih andal karena mampu mempelajari dependensi jangka panjang dan mengatasi permasalahan *vanishing gradient* [16], [17]. Beberapa penelitian menunjukkan bahwa LSTM memiliki kinerja lebih baik dibandingkan metode lain dalam prediksi jangka pendek iradiasi maupun daya PV [18], [19].

Selain pemilihan model, teknik *feature engineering* juga berperan penting dalam meningkatkan akurasi peramalan. Penggunaan *lag features* dari data historis, transformasi waktu siklikal berbasis fungsi sinus dan kosinus, serta klusterisasi kondisi cuaca dengan K-Means telah terbukti membantu model mengenali pola temporal secara lebih baik [21], [22]. Untuk mengoptimalkan performa model, pendekatan *Bayesian Optimization* digunakan dalam pemilihan *hyperparameter*, karena mampu mencari konfigurasi terbaik secara efisien pada ruang pencarian yang luas [23], [24].

3. METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan model prediksi iradiasi matahari harian per jam (*one-day-ahead hourly forecasting*) berbasis LSTM, dengan memanfaatkan data iradiasi aktual tanpa melibatkan variabel eksternal. Tahapan metode meliputi akuisisi data, pra-pemrosesan, rekayasa fitur, pembentukan data *time-series*, pelatihan model, serta evaluasi performa, yang tampak pada Gambar 1.



Gambar 1 Diagram Alir Metodologi untuk Pengembangan Model Peramalan Penyinaran Matahari Menggunakan LSTM

3.1. Persiapan Data

Pada tahap ini, data mentah diolah dan disesuaikan sehingga siap digunakan sebagai masukan untuk pelatihan model LSTM.

3.1.1. Akuisisi Data

Data iradiasi yang digunakan diperoleh dari platform Indonesia Solar Map, yang menyediakan data iradiasi matahari per jam berdasarkan hasil pengukuran menggunakan pyranometer [25]. Data yang tersedia berupa GHI dalam satuan watt per meter persegi (W/m^2), dicatat secara kontinu sepanjang hari, dari waktu cuplik 30 detik hingga 1 jam. Penelitian ini mengambil data GHI per jam lokasi Yogyakarta dari Januari 2023 hingga September 2023.

3.1.2. Pra-pemrosesan Data

Data yang mengandung nilai hilang (NaN) diisi menggunakan nilai rata-rata berdasarkan jam yang sama dalam seluruh dataset [26]. Sedangkan, data pada hari dimana tidak ada datanya, akan dibuang. Selanjutnya, data disaring untuk hanya mengambil periode antara pukul 06.00 hingga 18.00 (13 jam aktif per hari), sesuai rentang waktu penyinaran efektif.

Data iradiasi dinormalisasi menggunakan metode z-score untuk mengurangi bias skala dan mempercepat konvergensi saat pelatihan model [27]. Z-score mengubah nilai penyinaran matahari dengan rata-rata 0 dan standard deviasi 1 [28]. Hal ini dapat dicapai dengan menggunakan persamaan 1

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

Dimana x adalah nilai GHI, μ adalah rata-rata GHI, dan σ adalah standar deviasi GHI.

3.1.3. Transformasi Data & Rekayasa Fitur

Data GHI yang awalnya merupakan *timeseries* per jam dari Januari 2023 hingga September 2023 diubah menjadi bentuk per hari, dengan masing-masing sampel mencakup 13 jam data berturut-turut. Membagi data *time series* menjadi potongan-potongan urutan yang lebih kecil dapat secara signifikan membantu model LSTM mempelajari hubungan waktu di dalam data [29].

Data GHI harian dikelompokkan ke dalam tiga kluster menggunakan algoritma K-Means untuk menangkap variasi pola penyinaran. Karena tidak ada label pada data penyinaran matahari, digunakan pendekatan *unsupervised learning* [21]. Algoritma K-Means bekerja dengan meminimalkan *within-cluster sum of squares* (WCSS), yang merupakan fungsi objektif untuk menemukan partisi kluster yang optimal [30]. Fungsi objektif ini didefinisikan dalam persamaan 2

$$J = \sum_{i=1}^k \sum_{x \in S_i} ||x - \mu_i||^2 \quad (2)$$

Dimana J adalah total WCSS, k adalah jumlah kluster (dalam penelitian ini, 3) S_i adalah himpunan titik data di kluster ke- i , x adalah titik data individual, dan μ_i adalah titik pusat (*centroid*) dari kluster ke- i .

Setiap data harian kemudian diberi label numerik 1, 2, atau 3 untuk merepresentasikan kondisi penyinaran, yang secara umum dikategorikan sebagai cerah, berawan, dan gelap dimana label ini akan menjadi fitur dalam input LSTM.

Selain itu, label kluster dan nilai GHI dari tiga hari sebelumnya ($H-3$, $H-2$, $H-1$) digunakan sebagai fitur tambahan, menghasilkan tiga fitur waktu lampau (*lag features*). Informasi jam data GHI dikonversi menjadi dua fitur berbasis transformasi sinus dan kosinus untuk mempertahankan sifat siklikal waktu dalam input model dengan menerapkan persamaan 3 [22].

$$X_{sin} = \sin\left(\frac{2\pi h}{24}\right) \text{ dan } X_{cos} = \cos\left(\frac{2\pi h}{24}\right) \quad (3)$$

Dengan demikian, setiap titik waktu memiliki total 8 fitur input (3 fitur iradiasi historis + 3 fitur kluster + 2 fitur waktu).

3.1.4. Pemisahan Data

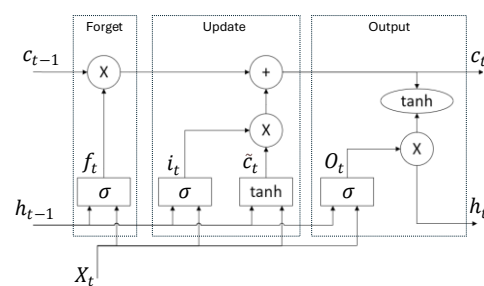
Data dibagi menjadi tiga bagian: pelatihan (training), validasi, dan pengujian (testing), dengan skema 64% untuk pelatihan 16% untuk validasi, dan 20% untuk pengujian

3.2. LSTM

LSTM adalah jenis khusus dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi masalah *vanishing gradient* sehingga mampu mempelajari dependensi jangka panjang antar data dalam sebuah sekuens [31].

Dalam penelitian ini, input untuk model LSTM adalah sebuah **sekuens (urutan) utuh** yang merepresentasikan satu hari pengamatan yaitu sebuah **matriks berukuran 8×13** , yang terdiri dari 8 fitur dengan 13 langkah waktu (jam 06:00 hingga 18:00) pada setiap fiturnya.

Model LSTM bekerja dengan "membaca" keseluruhan sekuens 13 jam ini secara berurutan, dari jam pertama hingga jam terakhir, untuk membangun pemahaman komprehensif tentang pola iradiasi pada hari tersebut. Proses ini diatur oleh mekanisme gerbang di dalam setiap unit sel memori LSTM, c_t .



Gambar 2 Struktur Unit Sel pada LSTM

Saat model bergerak dari satu jam ke jam berikutnya, pertama-tama *forget gate*, f_t , menentukan informasi mana dari sel memori sebelumnya (c_{t-1}) yang tidak lagi relevan dan harus dibuang. Keputusan ini didasarkan pada input saat ini (X_t) dan *hidden state* sebelumnya (h_{t-1}), yang memungkinkan model untuk membuang memori usang [29], seperti kondisi cerah di pagi hari ketika data sore menunjukkan perubahan cuaca menjadi mendung. Arsitektur model yang digunakan dalam penelitian ini memiliki

struktur satu lapisan tersembunyi (*single hidden layer*)

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (4)$$

Dimana W_f dan b_f adalah matriks bobot dan bias yang nilainya dipelajari oleh model selama pelatihan. Fungsi aktivasi σ memastikan outputnya bernilai antara 0 (“lupakan sepenuhnya”) dan 1 (“simpan sepenuhnya”).

Selanjutnya, *input gate*, i_t , memutuskan informasi baru mana dari input saat ini yang penting untuk disimpan. Gerbang ini, bersama dengan lapisan tanh yang menciptakan kandidat nilai baru $\tilde{c}_t$, memilih aspek-aspek yang paling signifikan dari 8 fitur yang masuk [31].

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (5)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \quad (6)$$

Sel memori kemudian diperbarui dengan menggabungkan memori lama yang relevan dengan informasi baru, c_t , yang telah disaring oleh gerbang masuk.

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (7)$$

Terakhir, *output gate*, O_t , menentukan output model (h_t) berdasarkan status sel yang telah diperbarui. Setelah seluruh sekuens 13 jam selesai diproses, *hidden state* terakhir berisi rangkuman informasi dari keseluruhan pola harian. Representasi ringkas inilah yang kemudian digunakan oleh lapisan selanjutnya untuk menghasilkan prediksi iradiasi matahari untuk satu hari ke depan [31].

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (8)$$

$$h_t = O_t \times \tanh(c_t) \quad (9)$$

3.3. Hyperparameter Tuning

Performa model prediksi berbasis *deep learning*, dan pembelajaran mesin secara umum, dapat diperbaiki melalui proses pengaturan *hyperparameter* yang tepat. Salah satu metode yang digunakan adalah Bayesian Optimization (BO), yaitu prosedur desain berurutan (*sequential design*) untuk optimasi global fungsi *black-box* yang tidak dapat diturunkan secara matematis [23]. BO membangun model probabilistik dari fungsi objektif untuk menentukan *hyperparameter* terbaik secara lebih terarah, sehingga

jumlah evaluasi fungsi dapat dikurangi dengan hanya menguji kombinasi yang paling menjanjikan.

Secara umum, BO memecahkan masalah dengan mencari parameter yang meminimalkan fungsi objektif dalam domain terbatas, di mana setiap variabel memiliki batas bawah dan batas atas, sebagaimana ditunjukkan pada Persamaan 10 [24].

$$x^* = \arg \min_{x \in X} f(x) \quad (10)$$

Dimana $f(x)$ adalah nilai yang akan diminimalkan, X adalah domain nilai *hyperparameter*, yang ditunjukkan pada Tabel 1, dan x^* merupakan kombinasi *hyperparameter* yang menghasilkan nilai terendah $f(x)$.

Dalam penelitian ini, nilai $f(x)$ yang diminimalkan adalah RMSE antara hasil prediksi LSTM dan data validasi. Proses optimasi dilakukan dengan menjalankan pelatihan model sebanyak 30 kali, kemudian dipilih kombinasi *hyperparameter* yang menghasilkan nilai RMSE terendah.

Tabel 1 Pemilihan dan rentang hyperparameter

Hyperparameter	Rentang	Type
Jumlah hidden unit LSTM	1 – 256	Integer
Laju dropout	0 – 0.4	Continuous
Ukuran batch LSTM	10 – jumlah hari	Integer
LSTM learning rate	0.00001 – 0.01	log

3.4. Evaluasi Model

Model yang telah dilatih dievaluasi menggunakan empat metrik regresi, yaitu Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R-Squared, R^2). RMSE mengukur perbedaan antara nilai aktual dan nilai hasil peramalan, di mana nilai yang lebih rendah menunjukkan hasil peramalan yang lebih baik. Secara matematis, RMSE didefinisikan sebagai

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_f - P_a)^2} \quad (11)$$

di mana N adalah jumlah pengamatan; P_f adalah nilai hasil peramalan; P_a adalah nilai aktual. MAE merupakan rata-rata dari kesalahan absolut antara

nilai hasil peramalan dan nilai aktual. Indikator ini mampu merefleksikan tingkat akurasi kesalahan peramalan dengan lebih baik. Rumusnya adalah

$$MAE = \frac{\sum_{i=1}^N |P_f - P_a|}{N} \quad (12)$$

MAPE menggambarkan rasio kesalahan terhadap nilai aktual dalam bentuk persentase, sehingga memudahkan interpretasi skala kesalahan. Persamaannya dinyatakan sebagai

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|P_f - P_a|}{P_a} \times 100\% \quad (13)$$

Sementara itu, R^2 digunakan untuk menilai tingkat kecocokan model (fitting) dan mengukur sejauh mana prediksi menyimpang dari kenyataan. Nilai R^2 berada pada rentang 0 hingga 1, di mana Jika $R^2 = 0$, menunjukkan kecocokan yang buruk dan $R^2 = 1$ menandakan prediksi tanpa kesalahan. Rumusnya adalah

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_f - P_a)^2}{\sum_{i=1}^N (\bar{P}_a - P_a)^2} \quad (14)$$

Keempat metrik ini digunakan secara bersamaan untuk menilai akurasi prediksi model dari berbagai sudut pandang, baik dalam bentuk absolut maupun persentase, serta untuk mengevaluasi kekuatan hubungan antara hasil prediksi dengan data aktual.

4. HASIL DAN PEMBAHASAN

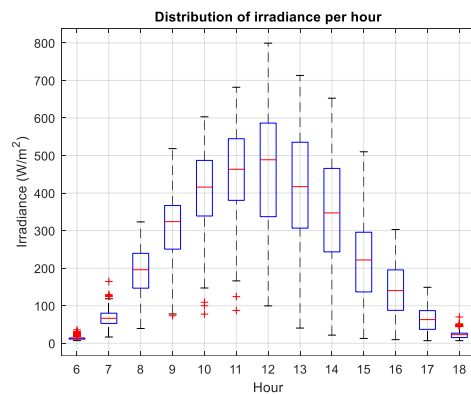
4.1. Karakteristik Dataset

Data penyinaran matahari pada penelitian ini dikumpulkan selama periode 255 hari, dengan pencatatan setiap jam mulai pukul 06.00 hingga 18.00. Tabel 2 menyajikan jumlah data yang hilang pada setiap jam pengamatan beserta persentasenya terhadap total hari observasi. Persentase data hilang bervariasi antara 26,7% hingga 47,8%, dengan jumlah terbesar terjadi pada pukul 06.00, yaitu 122 data hilang (47,8%). Informasi ini menjadi acuan dalam menentukan langkah pra-pemrosesan yang tepat sebelum analisis lanjutan dilakukan.

Jam	Jumlah Data Hilang	Jumlah Hari	% Hilang
6	122	255	47,8
7	79	255	31,0
8	76	255	29,8

Jam	Jumlah Data Hilang	Jumlah Hari	% Hilang
9	76	255	29,8
10	76	255	29,8
11	74	255	29,0
12	71	255	27,8
13	71	255	27,8
14	69	255	27,1
15	69	255	27,1
16	68	255	26,7
17	71	255	27,8
18	97	255	38,0

Setelah pra-pemrosesan, distribusi data iradiasi per jam ditunjukkan pada Gambar 3. Boxplot tersebut memperlihatkan pola penyinaran harian yang sesuai ekspektasi, di mana median iradiasi meningkat secara bertahap dari pagi hari hingga mencapai puncak sekitar pukul 12.00. Variabilitas data, yang ditunjukkan oleh rentang interkuartil (tinggi kotak), juga terlihat paling besar pada jam-jam tengah hari. Hal ini mengindikasikan fluktuasi kondisi cuaca yang lebih dinamis saat intensitas matahari sedang tinggi. Kehadiran beberapa data *outliers*, yang ditandai dengan simbol '+', menunjukkan adanya beberapa pembacaan anomali pada jam-jam tertentu.



Gambar 3 Distribusi penyinaran matahari tiap jam

Statistik deskriptif untuk masing-masing set data disajikan pada Tabel 3. Tabel tersebut menunjukkan bahwa ketiga set data—pelatihan, validasi, dan pengujian—memiliki karakteristik distribusi yang relatif serupa, terlihat dari nilai rata-rata (*mean*), median, dan standar deviasi yang sebanding. Kesamaan ini penting untuk memastikan bahwa model yang dilatih pada data pelatihan dapat dievaluasi secara adil dan mampu digeneralisasi dengan baik pada data pengujian yang belum pernah dilihat sebelumnya [32].

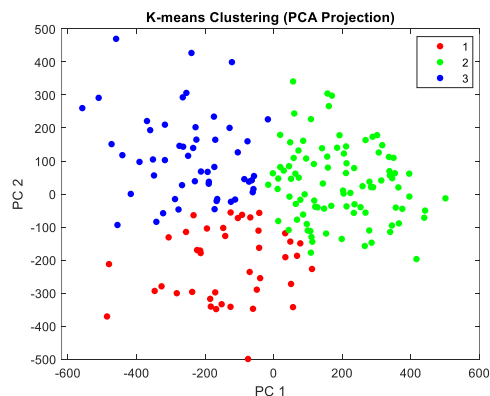
Tabel 3 Statistik Deskriptif Dataset

Dataset	Data Train	Data Validation n	Data Test
Jumlah Hari	117	30	37
Jumlah Data	1521	390	481
Rata-rata (W/m ²)	7	7	7
Median (W/m ²)	799,2	624,45	644,29
Std Dev (W/m ²)	233,5	231,92	264,48
Min (W/m ²)	206,73	212,79	263,45
Max (W/m ²)	189,65	176,19	190,23

4.2. Hasil Klasterisasi dan Tuning Model

Untuk mengidentifikasi variasi pola cuaca harian pada data iradiasi, penelitian ini menerapkan algoritma K-Means. Karena hasil klasterisasi K-Means menghasilkan data dalam ruang berdimensi tinggi, visualisasi pola klaster dilakukan menggunakan *Principal Component Analysis* (PCA). PCA adalah metode reduksi dimensi yang memproyeksikan data ke dalam sumbu utama (principal components) dengan mempertahankan sebagian besar variasi informasi, sehingga memudahkan pemetaan data ke dua dimensi [33].

Gambar 4 menunjukkan hasil visualisasi PCA yang membagi data ke dalam tiga klaster yang terpisah dengan jelas. Ketiga klaster tersebut merepresentasikan tiga kondisi cuaca dominan, yaitu cerah, berawan, dan sangat berawan. Label klaster ini kemudian dimanfaatkan sebagai fitur tambahan pada model LSTM untuk membantu model mengenali pola penyinaran yang berbeda pada setiap kondisi cuaca.



Gambar 4 Plot PCA Menunjukkan Tiga Klaster yang Diberi Label 1 (Merah), 2 (Hijau), dan 3 (Biru)

Selanjutnya, proses *tuning* hyperparameter dilakukan menggunakan *Bayesian Optimization* untuk menemukan konfigurasi model terbaik. Tabel 4 menyajikan kombinasi hyperparameter optimal yang berhasil meminimalkan nilai RMSE pada set validasi. Hyperparameter inilah yang kemudian digunakan untuk melatih model final.

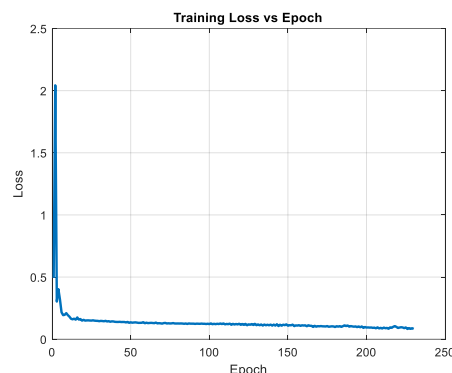
Tabel 4 Hyperparameter Optimal Model LSTM yang Dipilih Menggunakan Optimisasi Bayesian

Hyperparameter	Nilai Optimal
Jumlah Unit Tersembunyi	166
Laju Dropout	0.159
Ukuran Batch	56
Laju Pembelajaran	0.0090419

4.3. Evaluasi Performa Model Prediksi

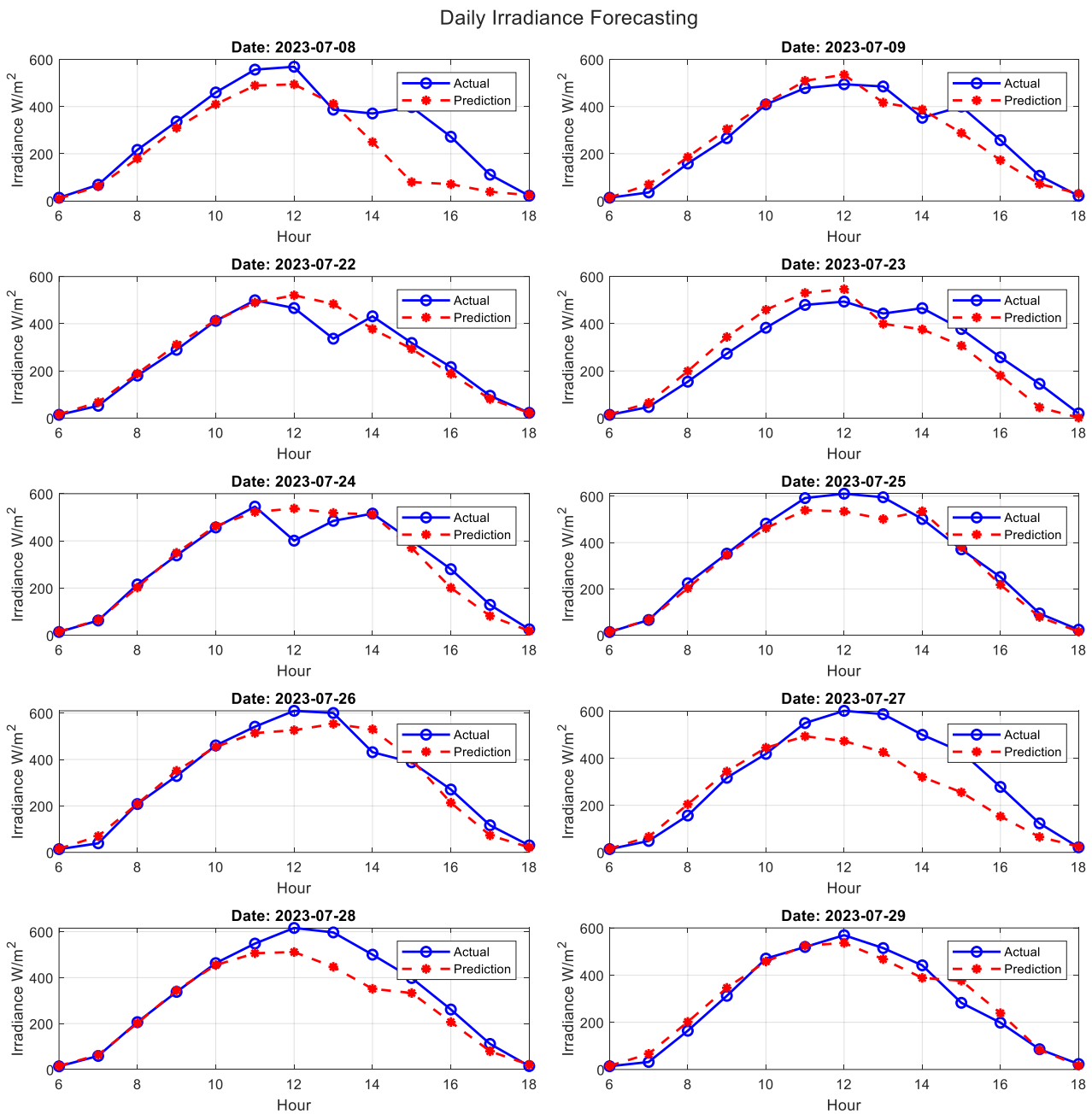
Untuk mengevaluasi kemampuan generalisasi model, performa model LSTM final diuji menggunakan set data pengujian yang belum pernah dilihat sebelumnya. Model yang diusulkan mencapai nilai **RMSE sebesar 87,71 W/m²** dan **MAE sebesar 59,34 W/m²**. Nilai **R-squared (R²) sebesar 0,7937** menjadi sorotan utama, yang mengindikasikan bahwa model mampu menjelaskan sekitar 79,4% variabilitas dalam data iradiasi matahari. Hal ini menunjukkan tingkat kecocokan yang kuat antara nilai prediksi dan nilai aktual. Sementara itu, nilai **MAPE sebesar 28,75%** merepresentasikan rata-rata persentase kesalahan dari prediksi.

Gambar 5 menunjukkan kurva *training loss* selama proses pelatihan, di mana terlihat bahwa nilai *loss* menurun secara konsisten dan konvergen, yang menandakan proses pelatihan berjalan dengan baik dan model berhasil mempelajari pola dari data.



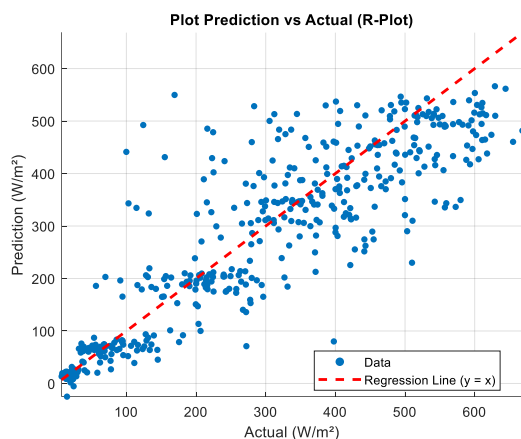
Gambar 5 Kurva *Training Loss* Model LSTM Selama 200 Epoch

Gambar 6 membandingkan nilai iradiasi aktual dengan hasil prediksi dari model LSTM. Terlihat bahwa model berhasil menangkap pola siklus harian—naik di pagi hari, puncak di siang hari, dan turun di sore hari—dengan cukup akurat. Meskipun ada beberapa deviasi pada saat terjadi fluktuasi tajam, secara umum alur prediksi mengikuti data aktual dengan baik.



Gambar 6 Perbandingan Iradiasi Surya Harian Aktual dan Hasil Prediksi untuk Beberapa Hari Menggunakan Model LSTM

Korelasi antara nilai prediksi dan aktual diperjelas lebih lanjut pada Gambar 7. *Scatter plot* ini menunjukkan bahwa titik-titik data cenderung mengelompok di sekitar garis diagonal ($y=x$), yang mengonfirmasi adanya hubungan linear yang kuat antara hasil prediksi dan data sebenarnya.



Gambar 7 Scatter Plot antara Prediksi dan Nilai Aktual Iradiasi Surya dengan Garis Regresi ($y = x$)

5. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi model LSTM untuk peramalan iradiasi matahari per jam satu hari ke depan dengan memanfaatkan data historis GHI sebagai satu-satunya input. Pendekatan ini memanfaatkan rekayasa fitur yang mencakup klusterisasi K-Means untuk identifikasi pola cuaca harian, pemanfaatan data historis beberapa hari sebelumnya, serta transformasi waktu siklikal, sehingga model mampu mengekstrak informasi temporal yang kompleks tanpa bergantung pada data meteorologi eksternal.

Hasil evaluasi pada testing set menunjukkan bahwa model mencapai nilai R-squared (R^2) sebesar 0,7937, yang berarti mampu menjelaskan sekitar 79,4% variabilitas pada data iradiasi aktual. Performa ini diperkuat oleh nilai RMSE sebesar 87,71 W/m^2 dan MAE sebesar 59,34 W/m^2 . Analisis visual melalui perbandingan grafik dan scatter plot mengonfirmasi bahwa model mampu mengikuti pola siklus harian dengan baik dan memiliki korelasi linear yang kuat terhadap data aktual.

Untuk penelitian selanjutnya, pengembangan dapat difokuskan pada transformasi data dari domain waktu ke representasi lain, seperti domain frekuensi atau

gabungan waktu–frekuensi. Pendekatan ini berpotensi mengekstraksi pola variabilitas iradiasi yang lebih jelas dan terstruktur, sehingga memudahkan model deep learning dalam mempelajari karakteristik data tanpa perlu menambahkan variabel eksternal.

DAFTAR PUSTAKA

- [1] Renewable Energy Policy Network for the 21st Century, “Renewables 2023 Global Status Report Collection,” REN21, 2023. [Online]. Available: <https://www.ren21.net/gsr-2023/>
- [2] IEA, “Renewables 2023,” IEA, Paris, 2024. [Online]. Available: <https://www.iea.org/reports/renewables-2023>
- [3] H. N. Amer, N. Y. Dahlan, A. M. Azmi, M. F. A. Latip, M. S. Onn, and A. Tumian, “Solar power prediction based on Artificial Neural Network guided by feature selection for Large-scale Solar Photovoltaic Plant,” *Energy Reports*, vol. 9, pp. 262–266, Nov. 2023, doi: 10.1016/j.egy.2023.09.141.
- [4] F. M. Guangul and G. T. Chala, “Solar Energy as Renewable Energy Source: SWOT Analysis,” in *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, Oman: IEEE, Jan. 2019, pp. 1–5. doi: 10.1109/ICBDSC.2019.8645580.
- [5] U. K. Das *et al.*, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, Jan. 2018, doi: 10.1016/j.rser.2017.08.017.
- [6] D. Rowe, S. Sayeef, and G. Platt, “Intermittency: It’s the Short-Term That Matters,” in *Future of Utilities Utilities of the Future*, Elsevier, 2016, pp. 129–150. doi: 10.1016/B978-0-12-804249-6.00007-5.
- [7] J. Ramirez-Vergara, L. B. Bosman, E. Wollega, and W. D. Leon-Salas, “Review of forecasting methods to support photovoltaic predictive maintenance,” *Cleaner Engineering and Technology*, vol. 8, p. 100460, June 2022, doi: 10.1016/j.clet.2022.100460.
- [8] S. S. Refaat, H. Abu-Rub, A. P. Sanfilippo, and A. Mohamed, “Impact of grid-tied large-scale photovoltaic system on dynamic voltage stability of electric power grids,” *IET Renewable Power Gen*, vol. 12, no. 2, pp. 157–164, Feb. 2018, doi: 10.1049/iet-rpg.2017.0219.
- [9] H. Liu, N. Nikitas, Y. Li, and R. Yang, “Big Data Analysis of Energy Economics in

- Photovoltaic Power Generation Market,” in *Big Data in Energy Economics*, in Management for Professionals. , Singapore: Springer Singapore, 2022, pp. 117–136. doi: 10.1007/978-981-16-8965-9_5.
- [10] K. Sereja and P. Miller, “Forecasting and evaluation of voltage conditions in a low-voltage power grid with a large share of prosumer installations,” in *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019*, R. S. Romaniuk and M. Linczuk, Eds., Wilga, Poland: SPIE, Nov. 2019, p. 188. doi: 10.1117/12.2536975.
- [11] N. Y. Hendrikx *et al.*, “All sky imaging-based short-term solar irradiance forecasting with Long Short-Term Memory networks,” *Sol. Energy*, vol. 272, 2024, doi: 10.1016/j.solener.2024.112463.
- [12] M. Ansong, T. N. Nyang’onda, R. J. Musembi, and B. S. Richards, “Very Short-Term Solar Irradiance Forecasting for Photovoltaic Power Integration with the Grid: Potentials and Challenges for Africa,” in *IEEE PES/IAS PowerAfrica, PowerAfrica*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/PowerAfrica61624.2024.10759405.
- [13] N. Omar, H. Aly, and T. Little, “LSTM and RBFNN Based Univariate and Multivariate Forecasting of Day-ahead Solar Irradiance for Atlantic Region in Canada and Mediterranean Region in Libya,” in *Int. Conf. Energy, Electr. Power Eng., CEEPE*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1130–1135. doi: 10.1109/CEEPE51765.2021.9475652.
- [14] M. AlShafeey and C. Csáki, “Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods,” *Energy Rep.*, vol. 7, pp. 7601–7614, 2021, doi: 10.1016/j.egy.2021.10.125.
- [15] H. Wang, P. Wang, Y. Zhang, and X. Sun, “Short-term Prediction of Solar Irradiance Based on EEMD-Lasso-MLP,” in *Proc. - Int. Conf. Mach. Learn. Knowl. Eng., MLKE*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 164–168. doi: 10.1109/MLKE55170.2022.00038.
- [16] S. J. Abdulkadir, H. Alhussian, M. Nazmi, and A. A. Elsheikh, “Long short term memory recurrent network for standard and Poor’s 500 index modelling,” *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 25–29, 2018, doi: 10.14419/ijet.v7i4.15.21365.
- [17] A. N. A. Mazmee, N. Zaini, and M. F. A. Latip, “Enhancing Solar Energy Forecasting Accuracy Using LSTM Networks for Global Horizontal Irradiance,” in *Int. Conf. Sys. Eng. Technol., ICSET - Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 174–179. doi: 10.1109/ICSET63729.2024.10774732.
- [18] P. S. Kumar, V. A. G. Raju, Ch. Ramesh, P. Prasad, T. Soniya, and A. Yaswanth, “Enhancement of Solar Energy Forecasting with LSTM Model: A Deep Dive into Accuracy and Performance Metrics,” in *IEEE Int. Conf. Adv. Inf. Technol., ICAIT - Proc.*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICAIT61638.2024.10690374.
- [19] R. A. Rajagukguk, R. A. A. Ramadhan, and H.-J. Lee, “A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power,” *Energies*, vol. 13, no. 24, 2020, doi: 10.3390/en13246623.
- [20] M. Lim and T. Handhayani, “PENERAPAN LSTM DAN GRU UNTUK PREDIKSI HARGA CABAI MERAH DI KOTA JAWA TIMUR,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6467.
- [21] W. Maseri Binti Wan Mohd, A. H. Beg, T. Herawan, A. Noraziah, and K. F. Rabbi, “Improved parameterless K-means: Auto-generation centroids and distance data point clusters,” in *Information Retrieval Methods for Multidisciplinary Applications*, Z. Lu, Ed., IGI Global, 2013, pp. 156–168. doi: 10.4018/978-1-4666-3898-3.ch010.
- [22] A. Zharova and A. Scherz, “Multistep Multiappliance Load Prediction,” Dec. 19, 2022, *arXiv: arXiv:2212.09426*. doi: 10.48550/arXiv.2212.09426.
- [23] N. Sultana, S. M. Z. Hossain, S. H. Almuhamini, and D. Düşteğör, “Bayesian Optimization Algorithm-Based Statistical and Machine Learning Approaches for Forecasting Short-Term Electricity Demand,” *Energies*, vol. 15, no. 9, p. 3425, Jan. 2022, doi: 10.3390/en15093425.
- [24] E. T. Habtemariam, K. Kekeba, M. Martínez-Ballesteros, and F. Martínez-Álvarez, “A Bayesian Optimization-Based LSTM Model for Wind Power Forecasting in the Adama District, Ethiopia,” *Energies*, vol. 16, no. 5, p. 2317, Jan. 2023, doi: 10.3390/en16052317.
- [25] AESI, “Indonesia Solar Map.” Accessed: Aug. 08, 2025. [Online]. Available: <https://indonesiasolarmap.com/methodology>
- [26] D. Pandya, A. Jadeja, S. Gour, S. B. Trivedi, H. H. Patel, and P. U. Jadeja, “An Analytical Perspective of Missing Values in Machine

- Learning,” in *Lect. Notes Networks Syst.*, Springer Science and Business Media Deutschland GmbH, 2024, pp. 285–294. doi: 10.1007/978-981-97-3991-2_24.
- [27] S. Urolagin, N. Sharma, and T. K. Datta, “A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting,” *Energy*, vol. 231, 2021, doi: 10.1016/j.energy.2021.120963.
- [28] A. Pranolo *et al.*, “Enhanced Multivariate Time Series Analysis Using LSTM: A Comparative Study of Min-Max and Z-Score Normalization Techniques,” *Ilk. J. Ilm.*, vol. 16, no. 2, pp. 210–220, Aug. 2024, doi: 10.33096/ilkom.v16i2.2333.210-220.
- [29] A. Katrompas and V. Metsis, “Enhancing LSTM Models with Self-attention and Stateful Training,” in *Lect. Notes Networks Syst.*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 217–235. doi: 10.1007/978-3-030-82193-7_14.
- [30] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [31] Y. Yu, J. Cao, and J. Zhu, “An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions,” *IEEE Access*, vol. 7, pp. 145651–145666, 2019, doi: 10.1109/ACCESS.2019.2946057.
- [32] J. Mielke, F. Innerbichler, M. Schiestl, N. M. Ballarini, and B. Jones, “The Assessment of Quality Attributes for Biosimilars: a Statistical Perspective on Current Practice and a Proposal,” *AAPS J.*, vol. 21, no. 1, 2019, doi: 10.1208/s12248-018-0275-9.
- [33] S. A. Mousavian Anaraki, A. Haeri, and F. Moslehi, “A hybrid reciprocal model of PCA and K-means with an innovative approach of considering sub-datasets for the improvement of K-means initialization and step-by-step labeling to create clusters with high interpretability,” *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1387–1402, 2021, doi: 10.1007/s10044-021-00977-x.