



## Hyperparameter Tuning Metode Long Short-Term Memory (LSTM) untuk Prediksi Harga Bawang Merah di Pasar Tradisional Jawa Timur

M. Latifur Rahman<sup>1</sup>, Mamluatul Hani'ah<sup>1\*</sup>, dan Muhammad Afif Hendrawan<sup>1</sup>

<sup>1</sup> Jurusan Teknologi Informasi, Politeknik Negeri Malang, Indonesia

\* Korespondensi: [mamluatulhaniah@polinema.ac.id](mailto:mamluatulhaniah@polinema.ac.id)

**Sitasi:** M. L. Rahman, M. Hani'ah, and M. A. Hendrawan, "Hyperparameter Tuning Metode Long Short-Term Memory (LSTM) untuk Prediksi Harga Bawang Merah di Pasar Tradisional Jawa Timur", *Jurnal Teknologi Informasi Dan Multimedia*, vol. 8, no. 2, pp. 417-430, 2026. <https://doi.org/10.35746/jtim.v8i2.976>

Diterima: 20-02-2026  
Direvisi: 24-05-2026  
Disetujui: tgl-05-2026



**Copyright:** © 2026 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

**Abstract:** Shallots are one of the essential food commodities in East Java, the price of which frequently fluctuates due to seasonal influences and distribution factors. This price uncertainty often weakens the community's purchasing power and makes it difficult for farmers and sellers to make decisions. Therefore, this study proposes a daily shallot price prediction model for nine regions in East Java using the Long Short-Term Memory (LSTM) method with hyperparameter tuning to obtain the best-performing model. This study utilizes daily price data from PIHPS (Pusat Informasi Harga Pangan Strategis / Strategic Food Price Information Center) over four years, from January 1, 2021, to December 31, 2024. The research stages begin with data preprocessing, which includes imputing missing values to ensure continuity of the time sequence, data normalization, and arrangement of data based on temporal order (time series). To obtain the optimal model, the data was split into 80% for training and 20% for testing. Based on the experimental results, the optimal model configuration was found using a Window Size of 7 (the last seven days of data), 114 Epochs, and a Batch Size of 64. Furthermore, the reliability of the model was measured using MAPE, where the average across all regions fell below 10%, which is classified as "Very Good" based on MAPE interpretation standards. Additional testing using 2025 data as an independent dataset further confirmed that the model remains consistent and stable, as evidenced by an average MAPE of 3.10% across the nine regions.

**Keywords:** Red Onion Price Prediction, Time Series Data, Long Short-Term Memory (LSTM), MAPE, Jawa Timur

**Abstrak:** Bawang merah adalah salah satu kebutuhan pangan penting di Jawa Timur yang harganya sering kali naik-turun (fluktuatif) akibat pengaruh musim dan distribusi. Ketidakpastian harga ini sering memicu lemahnya daya beli masyarakat dan menyulitkan petani serta pedagang dalam mengambil keputusan. Oleh karena itu, penelitian ini mengusulkan prediksi harga harian bawang merah di sembilan wilayah Jawa Timur menggunakan metode *Long Short-Term Memory* (LSTM) dengan *hyperparameter tuning* untuk mendapatkan model terbaik. Penelitian ini menggunakan data harga harian dari PIHPS (Pusat Informasi Harga Pangan Strategis) selama empat tahun, mulai 1 Januari 2021 hingga 31 Desember 2024. Tahapan penelitian dimulai dari data *preprocessing* yang meliputi pengisian data kosong agar urutan waktu tidak terputus, penyeragaman skala data (*normalization*), hingga penyusunan data berdasarkan urutan waktu (*time series*). Untuk mendapatkan model terbaik, model dilatih dengan pembagian data 80% untuk data latih dan 20% untuk data uji. Berdasarkan hasil pengujian, konfigurasi model optimal ditemukan pada penggunaan *Window Size* 7 (data seminggu terakhir), *Epoch* 114, dan *Batch Size* 64. Selanjutnya,

keandalan model diukur dengan menggunakan MAPE, di mana rata-rata di seluruh wilayah berada di bawah 10%, yang masuk dalam kategori "Sangat Baik" berdasarkan interpretasi MAPE. Pengujian tambahan menggunakan data tahun 2025 sebagai data independen juga membuktikan bahwa model ini tetap konsisten dan stabil, dibuktikan dengan rata-rata MAPE pada sembilan wilayah yaitu 3.10%.

**Kata kunci:** Prediksi Harga Bawang Merah, Data Time Series, *Long Short-Term Memory* (LSTM), MAPE, Jawa Timur

## 1. Pendahuluan

Jawa Timur memegang peran strategis dalam perekonomian nasional, sebagai salah satu sentra produksi pangan utama maupun wilayah dengan populasi penduduk yang besar. Salah satu komoditas yang memiliki permintaan tinggi sepanjang tahun di wilayah ini adalah bawang merah. Namun, komoditas ini menghadapi tantangan naik turunnya harga akibat pengaruh faktor musiman, perubahan cuaca yang tidak menentu, kendala distribusi, hingga lonjakan permintaan pada momen hari besar keagamaan [1]. Data Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) periode 2021–2024 merekam fluktuasi harga yang signifikan di pasar tradisional, dengan rentang terendah Rp18.950 hingga tertinggi Rp71.300 per kilogram[2]. Ketidakstabilan harga ini bukan hanya masalah statistik, melainkan berkontribusi langsung terhadap daya beli masyarakat dan menciptakan ketidakpastian pendapatan bagi petani [3].

Berbagai pendekatan metode peramalan telah diterapkan dalam penelitian terdahulu untuk mengatasi masalah ini, namun dengan performa yang beragam. Pendekatan statistik seperti yang dilakukan oleh [4] menggunakan *Seasonal ARIMA* (SARIMA) terbukti cukup efektif menangkap pola musiman produksi bawang merah dengan *Mean Absolute Percentage Error* (MAPE) sebesar 2.01%[4]. Namun, metode statistik ini sering kali kurang responsif terhadap perubahan data yang dinamis. Di sisi lain, pendekatan yang lebih sederhana seperti Regresi Linear yang diterapkan oleh [5] menghasilkan akurasi yang kurang optimal dengan MAPE sebesar 21.7%, menunjukkan keterbatasan metode linier dalam menangani data harga yang memiliki fluktuasi tajam [5]. Selain pendekatan tersebut, algoritma *Machine Learning* seperti *Support Vector Machine* (SVM) sering digunakan untuk permasalahan prediksi. Namun, algoritma tersebut sering kali memiliki performa yang kurang optimal untuk menangani data runtun waktu yang panjang. Hal ini sejalan dengan penelitian [6], yang membuktikan bahwa performa SVM kurang unggul dibandingkan dengan metode yang menggunakan *Deep Learning* seperti LSTM, LSTM terbukti lebih akurat dibandingkan SVM dalam kasus prediksi pada data *time series* [6]. Penelitian Agus Triyadi yang menerapkan *Long Short-Term Memory* (LSTM) untuk prediksi harga bawang merah di Wonosobo berhasil mencapai akurasi tinggi dengan MAPE 1,80% pada data validasi [7].

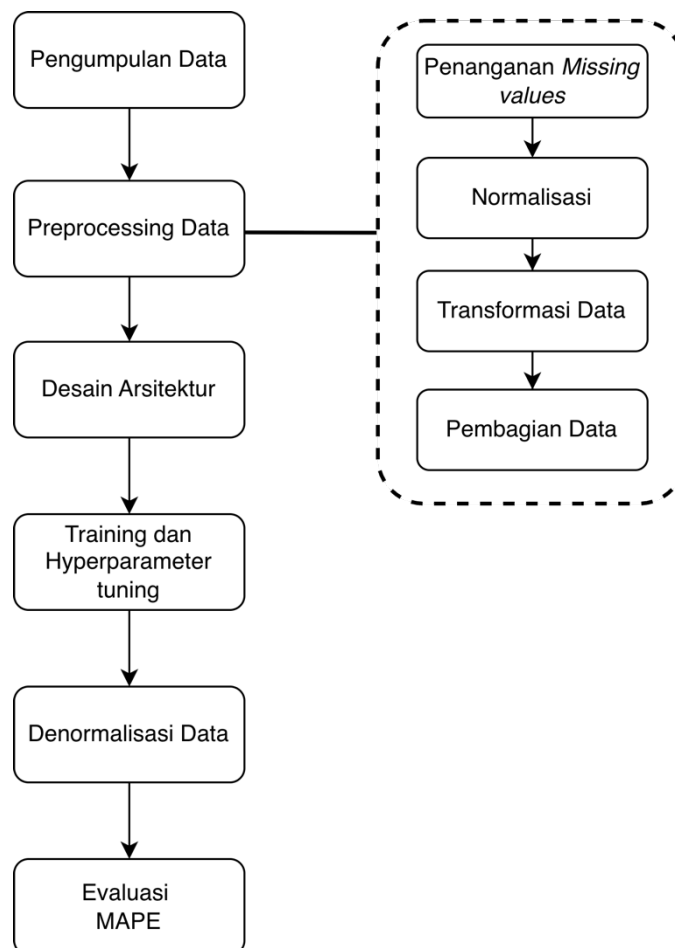
Meskipun LSTM menunjukkan kinerja yang unggul, penelitian-penelitian sebelumnya cenderung terbatas pada ruang lingkup satu lokasi atau pasar tertentu saja seperti pada penelitian [7] yang hanya dilakukan di Wonosobo. Belum banyak penelitian yang menggali performa model LSTM ketika diterapkan secara serentak di berbagai titik pasar dengan karakteristik data yang bervariasi dalam satu provinsi sentra produksi. Jawa Timur memiliki karakteristik pasar yang berbeda antar kabupaten/kota, sehingga model yang hanya diuji di satu titik belum tentu andal untuk wilayah lain. Selain itu, pada penelitian sebelumnya belum dibahas terkait *hyperparameter* yang optimal untuk model LSTM pada data harga bawang merah. Oleh karena itu, penelitian ini mengusulkan *hyperparameter tuning* metode *Long Short-Term Memory* (LSTM) untuk memprediksi harga harian bawang merah di sembilan wilayah strategis di Jawa Timur. Pendekatan ini merupakan pembaruan dari penelitian terdahulu karena tidak hanya berfokus pada

pencapaian *error* yang kecil di satu wilayah saja, melainkan menguji stabilitas model pada skala regional menggunakan dataset (2021-2024) pada sembilan wilayah berbeda.

LSTM dipilih karena merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengatasi masalah *vanishing gradient* [8], sehingga mampu memodelkan data sekuensial (runtun waktu) dan menangkap pola ketergantungan jangka panjang (*long-term dependencies*) maupun pola non-linear yang kompleks [9]. Penelitian ini bertujuan untuk mengidentifikasi kombinasi *hyperparameter* LSTM yang optimal pada data harga bawang merah lintas wilayah serta mengevaluasi kemampuan generalisasi model pada sembilan pasar dengan karakteristik yang berbeda.

## 2. Metodologi Penelitian

Untuk mencapai hasil prediksi harga yang optimal, terdapat serangkaian tahap penelitian yang harus dilakukan secara berurutan. Tahap-tahap ini dapat dilihat pada Gambar 1.



**Gambar. 1** Tahapan Penelitian.

### 2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa runtun waktu (*time series*) harga harian bawang merah di tingkat konsumen (pasar tradisional). Data diperoleh dari basis data publik Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) yang dikelola oleh Bank Indonesia [1]. Dataset mencakup periode 1 Januari 2021 hingga 31 Desember 2024 untuk 9 wilayah di Jawa Timur, yaitu Kota Surabaya, Malang, Kediri, Madiun, Probolinggo, Blitar, Kabupaten Jember, Banyuwangi, dan Sumenep. Pada tabel 1 merupakan data set dari 9 wilayah di Jawa Timur.

### 2.2. Preprocessing

Tahapan pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum masuk ke tahap pelatihan model.

#### 1. Penanganan *Missing Values*

Data kosong yang disebabkan oleh hari libur nasional ditangani menggunakan metode *Forward Fill*, yaitu dengan mengganti nilai kosong menggunakan data valid terakhir untuk menjaga kesinambungan pola data time series[10].

#### 2. Normalisasi

Data harga ditransformasi ke dalam rentang [0, 1] menggunakan metode *Min-Max Scaling* [11]. Langkah ini dilakukan untuk mempercepat proses konvergensi model karena LSTM sensitif terhadap skala data yang besar (ribuan Rupiah). Rumus *Min-Max Scaling* ditunjukkan dalam Persamaan (1) berikut [12]:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{1}$$

**Tabel. 1** Dataset 9 Wilayah di Jawa Timur

| N<br>o | Kota        | 10/02/2<br>021 | 11/02/2<br>021 | 12/02/2<br>021 | 15/02/2<br>021 | 16/02/2<br>021 | 17/02/2<br>021 | 18/02/2<br>021 | 19/02/2<br>021 | 22/02/<br>2021 |
|--------|-------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1      | Surabaya    | 26,750         | 26,750         | -              | 28,750         | 29,750         | 29,750         | 31,000         | 31,000         | 28,750         |
| 2      | Malang      | 26,500         | 26,250         | -              | 30,250         | 30,250         | 30,250         | 30,500         | 30,000         | 31,750         |
| 3      | Kediri      | 25,250         | 26,750         | -              | 27,750         | 28,500         | 28,500         | 28,500         | 28,500         | 31,000         |
| 4      | Jember      | 28,650         | 28,650         | -              | 30,000         | 30,650         | 30,650         | 30,000         | 30,350         | 30,650         |
| 5      | Banyuwangi  | 25,750         | 25,750         | -              | 27,750         | 30,500         | 30,500         | 29,500         | 29,500         | 29,250         |
| 6      | Madiun      | 24,000         | 24,000         | -              | 31,250         | 31,250         | 31,250         | 31,250         | 31,250         | 32,000         |
| 7      | Probolinggo | 20,500         | 21,750         | -              | 22,500         | 23,000         | 25,000         | 24,000         | 24,000         | 27,750         |
| 8      | Sukamenep   | 24,250         | 27,500         | -              | 27,750         | 27,250         | 27,250         | 27,000         | 27,000         | 30,250         |
| 9      | Blitar      | 22,750         | 22,750         | -              | 24,000         | 25,750         | 26,250         | 26,250         | 26,000         | 25,000         |

#### 3. Transformasi data

Transformasi data menggunakan teknik *sliding window* yaitu dengan menggeser *window* secara bertahap sesuai dengan ukuran *window* yang telah ditentukan [3]. Metode *sliding window* memecah sinyal kontinu menjadi segmen-segmen berdurasi tetap untuk memungkinkan analisis data secara berkelanjutan (*real-time*) [13]. Tabel 2 di bawah ini merupakan contoh dari penerapan *sliding window*

**Tabel. 2** Sampel Data Hasil Transformasi *Sliding Window*

| H-7    | H-6    | H-5           | H-4           | H-3           | H-2           | H-1           | Target<br>(H+1) |
|--------|--------|---------------|---------------|---------------|---------------|---------------|-----------------|
| 28,250 | 28,250 | 28,250        | 28,250        | 28,250        | 28,250        | 27,250        | <b>27,250</b>   |
| 28,250 | 28,250 | 28,250        | 28,250        | 28,250        | 27,250        | <b>27,250</b> | 27,250          |
| 28,250 | 28,250 | 28,250        | 28,250        | 27,250        | <b>27,250</b> | 27,250        | 27,750          |
| 28,250 | 28,250 | 28,250        | 27,250        | <b>27,250</b> | 27,250        | 27,750        | 27,250          |
| 28,250 | 28,250 | 27,250        | <b>27,250</b> | 27,250        | 27,750        | 27,250        | 27,750          |
| 28,250 | 27,250 | <b>27,250</b> | 27,250        | 27,750        | 27,250        | 27,750        | 27,750          |

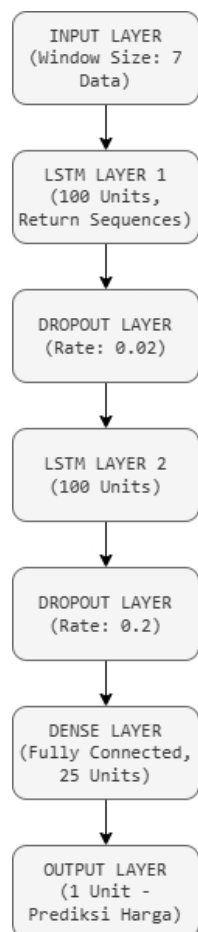
| H-7    | H-6    | H-5    | H-4    | H-3    | H-2    | H-1    | Target (H+1) |
|--------|--------|--------|--------|--------|--------|--------|--------------|
| 27,250 | 27,250 | 27,250 | 27,750 | 27,250 | 27,750 | 27,750 | 27,250       |
| 27,250 | 26,250 | 27,750 | 26,250 | 27,750 | 27,750 | 26,250 | 26,250       |

#### 4. Pembagian Data

Dataset yang telah ditransformasi selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20. Pembagian 80% data awal digunakan sebagai data latih (training) dan 20% data akhir sebagai data uji (testing).

#### 2.3. Desain Arsitektur dan Hyperparameter

Struktur model diawali dengan *Input Layer* yang menerima data dalam format *sliding window* berukuran 7 hari. Arsitektur ini menggunakan 2 lapis LSTM dengan masing-masing 100 *unit*. Lapisan pertama dikonfigurasi dengan *return sequences* untuk meneruskan *output* urutan ke lapisan berikutnya. Untuk mencegah *overfitting*, diterapkan *Dropout Layer* dengan *rate* sebesar 0.02 setelah setiap lapisan LSTM. Tahap pemrosesan akhir dilakukan oleh *Dense Layer* dengan 25 *unit*. Pada tahapan ini, *hyperparameter* yang digunakan adalah *optimizer* Adam dengan *learning rate* sebesar 0.001. Selanjutnya, *hyperparameter* akan disesuaikan saat pelatihan. Struktur detail arsitektur dapat dilihat pada Gambar 2.



**Gambar. 2** Konfigurasi Arsitektur Model

#### 2.4. Pelatihan model

Pelatihan melibatkan iterasi menggunakan data latih sebesar 80% dari dataset yang sudah dibagi sebelumnya, penyesuaian bobot berdasarkan perbedaan antara prediksi dan

nilai sebenarnya. Pemilihan *hyperparameter* berupa *optimizer*, jumlah *epoch*, dan *batch size* memiliki peran penting karena akan berdampak pada performa model. Setelah pelatihan, model dapat digunakan untuk membuat prediksi pada data baru dan evaluasi dilakukan untuk mengukur kinerja dan kemampuan generalisasi model [14]. Pada penelitian ini pelatihan menerapkan *hyperparameter* berupa 114 *epoch* dan *batch size* sebesar 64. Penggunaan 114 *epoch* berfungsi untuk memproses keseluruhan data latih sebanyak 114 kali putaran untuk mencapai *convergence*, sementara *batch size* 64 menentukan jumlah sampel yang diproses sebelum bobot model diperbarui. Setelah proses pelatihan selesai, model divalidasi menggunakan data uji dan *independent dataset*.

2.5. Denormalisasi Data

Denormalisasi data dilakukan untuk mengembalikan nilai hasil prediksi kedalam rentang nilai aslinya untuk mendapatkan nilai yang diharapkan dan dievaluasi dari model hasil evaluasi. Data asli yang telah dilakukan normalisasi memiliki rentang nilai 0-1 sehingga perlu dikembalikan ke data rentang aslinya untuk dapat dihitung hasil prediksinya [15]. Rumus denormalisasi pada range [0, 1] ditunjukkan pada persamaan (2) [16].

$$x = x'(max - min) + min \tag{2}$$

Keterangan:

- x = nilai denormalisasi
- x' = nilai prediksi yang masih dinormalisasi
- max = nilai maksimum
- min = nilai minimum

2.6. Evaluasi model

Kinerja model dievaluasi menggunakan metrik *Mean Absolute Percentage Error* (MAPE) untuk mengukur rata-rata persentase kesalahan absolut antara nilai prediksi dan nilai aktual. Perhitungan MAPE dirumuskan dalam Persamaan (3) [17].

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{y_n - \hat{Y}_n}{y_n} \right| \times 100\% \tag{3}$$

Dimana:

- MAPE = Mean Absolute Percentage Error
- N = Jumlah data pengamatan
- $y_n$  = Nilai sebenarnya dari data ke-n
- $\hat{Y}_n$  = Nilai prediksi ke-n
- |·| = Nilai absolut dari selisih

Selanjutnya, nilai MAPE yang diperoleh diklasifikasikan ke dalam empat kategori tingkat akurasi untuk menilai keandalan model, sebagaimana ditunjukkan pada Tabel 3.

**Tabel. 3** Range Nilai MAPE (Sumber: [18])

| Range MAPE (Mean Absolute Percentage Error) | Keterangan (Kompetensi Model Peramalan) |
|---|---|
| < 10 %                                      | Sangat Baik                             |
| 10 - 20 %                                   | Baik                                    |
| 20 - 50 %                                   | Layak                                   |
| > 50%                                       | Buruk                                   |

### 3. Hasil dan Pembahasan

Sebelum dilakukan pelatihan model, data mentah dari PIHPS yang mencakup periode 2021–2024 telah melalui tahapan *preprocessing* untuk memperbaiki data agar sesuai dengan masukan yang dibutuhkan oleh LSTM. Hasil dari tahap *preprocessing* menunjukkan bahwa format data yang semula melebar dalam bentuk kolom berhasil diubah menjadi format memanjang dalam bentuk baris dan dikelompokkan menjadi sembilan dataset terpisah sesuai wilayahnya. Pada tabel 4-6 merupakan hasil dari 3 contoh wilayah yang telah dikelompokkan.

**Tabel. 4** Dataset Kota Surabaya

| No | Kota     | Tanggal    | Harga  |
|----|----------|------------|--------|
| 1  | Surabaya | 10/02/2021 | 26,750 |
| 2  | Surabaya | 11/02/2021 | 26,750 |
| 3  | Surabaya | 12/02/2021 | -      |
| 4  | Surabaya | 15/02/2021 | 28,750 |
| 5  | Surabaya | 16/02/2021 | 29,750 |
| 6  | Surabaya | 17/02/2021 | 29,750 |
| 7  | Surabaya | 18/02/2021 | 31,000 |
| 8  | Surabaya | 19/02/2021 | 31,000 |
| 9  | Surabaya | 22/02/2021 | 28,750 |

**Tabel. 5** Dataset Kota Malang

| No | Kota   | Tanggal    | Harga  |
|----|--------|------------|--------|
| 1  | Malang | 10/02/2021 | 26,500 |
| 2  | Malang | 11/02/2021 | 26,250 |
| 3  | Malang | 12/02/2021 | -      |
| 4  | Malang | 15/02/2021 | 30,250 |
| 5  | Malang | 16/02/2021 | 30,250 |
| 6  | Malang | 17/02/2021 | 30,250 |
| 7  | Malang | 18/02/2021 | 30,500 |
| 8  | Malang | 19/02/2021 | 30,000 |
| 9  | Malang | 22/02/2021 | 31,750 |

**Tabel. 6** Dataset Kota Kediri

| No | Kota   | Tanggal    | Harga  |
|----|--------|------------|--------|
| 1  | Kediri | 10/02/2021 | 25,250 |
| 2  | Kediri | 11/02/2021 | 26,750 |
| 3  | Kediri | 12/02/2021 | -      |
| 4  | Kediri | 15/02/2021 | 27,750 |
| 5  | Kediri | 16/02/2021 | 28,500 |
| 6  | Kediri | 17/02/2021 | 28,500 |
| 7  | Kediri | 18/02/2021 | 28,500 |
| 8  | Kediri | 19/02/2021 | 28,500 |
| 9  | Kediri | 22/02/2021 | 31,000 |

Setelah data dikelompokkan per wilayah masing-masing, kemudian dilakukan identifikasi untuk menemukan adanya *missing values* pada hari libur nasional yang kemudian ditangani menggunakan metode *Forward Fill* sehingga data akan kontinu. Tabel di bawah ini merupakan contoh hasil penanganan *missing values* menggunakan metode *Forward Fill* pada tiga kota berbeda-beda.

Tabel. 7 Dataset Kota Surabaya

| No | Kota     | Tanggal    | Harga  |
|----|----------|------------|--------|
| 1  | Surabaya | 10/02/2021 | 26,750 |
| 2  | Surabaya | 11/02/2021 | 26,750 |
| 3  | Surabaya | 12/02/2021 | 26,750 |
| 4  | Surabaya | 15/02/2021 | 28,750 |
| 5  | Surabaya | 16/02/2021 | 29,750 |
| 6  | Surabaya | 17/02/2021 | 29,750 |
| 7  | Surabaya | 18/02/2021 | 31,000 |
| 8  | Surabaya | 19/02/2021 | 31,000 |
| 9  | Surabaya | 22/02/2021 | 28,750 |

Tabel. 8 Dataset Kota Malang

| No | Kota   | Tanggal    | Harga  |
|----|--------|------------|--------|
| 1  | Malang | 10/02/2021 | 26,500 |
| 2  | Malang | 11/02/2021 | 26,250 |
| 3  | Malang | 12/02/2021 | 26,250 |
| 4  | Malang | 15/02/2021 | 30,250 |
| 5  | Malang | 16/02/2021 | 30,250 |
| 6  | Malang | 17/02/2021 | 30,250 |
| 7  | Malang | 18/02/2021 | 30,500 |
| 8  | Malang | 19/02/2021 | 30,000 |
| 9  | Malang | 22/02/2021 | 31,750 |

Tabel. 9 Dataset Kota Kediri

| No | Kota   | Tanggal    | Harga  |
|----|--------|------------|--------|
| 1  | Kediri | 10/02/2021 | 25,250 |
| 2  | Kediri | 11/02/2021 | 26,750 |
| 3  | Kediri | 12/02/2021 | 26,750 |
| 4  | Kediri | 15/02/2021 | 27,750 |
| 5  | Kediri | 16/02/2021 | 28,500 |
| 6  | Kediri | 17/02/2021 | 28,500 |
| 7  | Kediri | 18/02/2021 | 28,500 |
| 8  | Kediri | 19/02/2021 | 28,500 |
| 9  | Kediri | 22/02/2021 | 31,000 |

Jumlah data disetiap wilayah adalah 1.043 baris sebelum dilakukan perbaikan *missing values*, sedangkan data untuk 9 wilayah berjumlah 9.387. Ditemukan *missing values* yang bervariasi antara 27 hingga 40 data kosong per wilayah, dengan total akumulasi 292 *missing values* dari keseluruhan dataset. Tidak ada data yang dihapus dalam tahapan ini. Seluruh *missing value* sdilengkapi menggunakan metode *Forward Fill* sehingga jumlah data tetap utuh sebanyak 1.043 baris valid per wilayah. Selain itu, data berupa lonjakan harga ekstrem tetap dipertahankan dan tidak dibuang, karena lonjakan tersebut adalah representasi riil dari dinamika pasar yang harus dipelajari polanya oleh model LSTM.

Selanjutnya, proses normalisasi dengan *Min-Max Scaling* mengonversi seluruh data harga nominal ke dalam rentang  $[0, 1]$ , yang diikuti dengan transformasi menggunakan teknik *Sliding Window* dengan *window size* 7 untuk membentuk struktur data 3 dimensi (*Samples, Time Steps, Features*) sebagai input model LSTM. Tabel 10 merupakan data yang telah dinormalisasi dengan menggunakan rumus *Min-Max Scaling*

Tabel. 10 Dataset hasil normalisasi

| No | Tanggal    | Harga  | Harga Normalisasi |
|----|------------|--------|-------------------|
| 1  | 10/02/2021 | 25,250 | 0.000             |
| 2  | 11/02/2021 | 26,750 | 0.261             |

| No | Tanggal    | Harga  | Harga Normalisasi |
|----|------------|--------|-------------------|
| 3  | 12/02/2021 | 26,750 | 0.261             |
| 4  | 15/02/2021 | 27,750 | 0.435             |
| 5  | 16/02/2021 | 28,500 | 0.565             |
| 6  | 17/02/2021 | 28,500 | 0.565             |
| 7  | 18/02/2021 | 28,500 | 0.565             |
| 8  | 19/02/2021 | 28,500 | 0.565             |
| 9  | 22/02/2021 | 31,000 | 1.000             |

Setelah tahapan normalisasi data dan transformasi menggunakan teknik *Sliding Window*, data dibagi menjadi 2, yaitu data latih dan data uji. Pengujian untuk menentukan konfigurasi *hyperparameter* terbaik menggunakan 3 skenario. Skenario pertama digunakan untuk mencari *window size* terbaik, skenario kedua untuk mencari *epoch* terbaik, dan skenario ketiga untuk mencari *batch size* terbaik. Hasil dari uji coba mencari konfigurasi terbaik dapat dilihat pada Tabel 11 hingga Tabel 13. Tabel 11 menunjukkan bahwa uji coba menggunakan *window size* 7 menghasilkan nilai MAPE paling rendah, yaitu **3.1441%**. Maka dari itu, *window size* 7 dipilih menjadi konfigurasi *window size* terbaik.

Tabel 12 menunjukkan bahwa uji coba menggunakan *batch size* 64 menghasilkan nilai MAPE paling rendah, yaitu **3.3422%**. Maka dari itu, *batch size* 64 dipilih menjadi konfigurasi *batch size* terbaik. Berdasarkan hasil uji coba pada Tabel 13, didapatkan nilai MAPE yang paling rendah dengan menggunakan *epoch* 114, yaitu pada *window size* 7, dengan nilai MAPE sebesar **2.6390%**. Setelah dilakukan serangkaian uji coba untuk mencari konfigurasi terbaik yang diuraikan sebelumnya, ditemukan kombinasi final yang memberikan nilai error terendah, yakni **Window Size 7, Epoch 114, Batch Size 64, dan Dropout 0.02**, yang kemudian diterapkan secara konsisten untuk 9 wilayah pengujian.

**Tabel. 11** Hasil Uji Coba Dengan Nilai *Window Size* Terbaik

| No | Skenario | <i>Window Size</i> | <i>Epoch</i> | <i>Batch Size</i> | <i>Dropout</i> | MAPE (%) |
|----|----------|--------------------|--------------|-------------------|----------------|----------|
| 1  | W1       | 7                  | 50           | 64                | 0.02           | 3.1441%  |
| 2  | W2       | 14                 | 50           | 64                | 0.02           | 4.6119%  |
| 3  | W3       | 90                 | 50           | 64                | 0.02           | 3.3422%  |

**Tabel. 12** Hasil Dari Mencari Konfigurasi *Batch Size* Terbaik

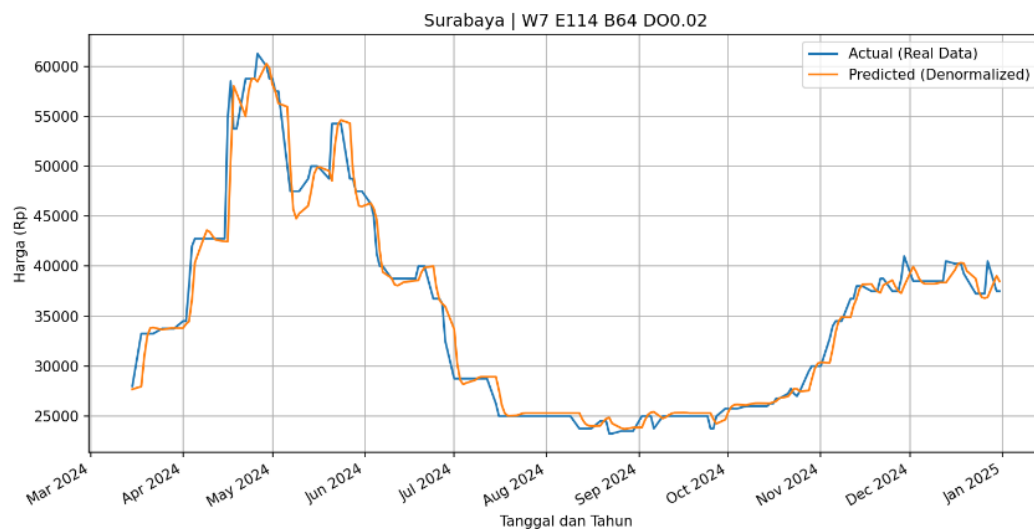
| No | Skenario | <i>Window Size</i> | <i>Epoch</i> | <i>Batch Size</i> | MAPE (%) |
|----|----------|--------------------|--------------|-------------------|----------|
| 1  | C1       | 90                 | 50           | 16                | 5.5420%  |
| 2  | C2       | 90                 | 50           | 30                | 3.3472%  |
| 3  | C3       | 90                 | 50           | 64                | 3.3422%  |

**Tabel. 13** Hasil Uji Coba Meningkatkan Nilai *Epoch*

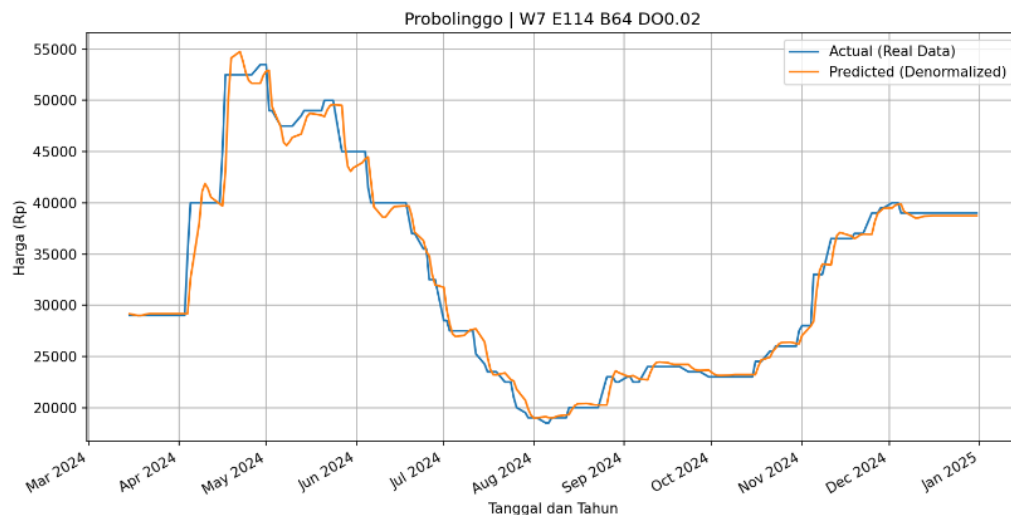
| No | Skenario | <i>Window Size</i> | <i>Epoch</i> | <i>Batch Size</i> | <i>Dropout</i> | MAPE (%) |
|----|----------|--------------------|--------------|-------------------|----------------|----------|
| 1  | E7-0.02  | 7                  | 114          | 64                | 0.02           | 2.6390%  |
| 2  | E90-0.02 | 90                 | 114          | 64                | 0.02           | 3.1775%  |

Selanjutnya, untuk mengetahui performa prediksi, evaluasi visual dilakukan untuk melihat kemampuan model dalam mengikuti pola data aktual vs data uji, di mana data uji merupakan data 20% yang sudah dibagi sebelumnya. Gambar 3 dan Gambar 4 menggambarkan perbandingan antara harga aktual (garis biru) dan hasil prediksi model LSTM (garis oranye) pada dua wilayah dengan karakteristik berbeda, yaitu Kota Surabaya dan

Kota Probolinggo. Meskipun terdapat fluktuasi tajam dan lonjakan harga ekstrem pada pertengahan tahun, model terbukti responsif dalam mengikuti arah pergerakan tren tersebut. Hal ini menunjukkan bahwa konfigurasi *Window Size 7* cukup efektif dalam menangkap pola pergerakan harga.



**Gambar 3** Grafik Perbandingan Harga Aktual vs Prediksi Kota Surabaya



**Gambar 4** Grafik Perbandingan Harga Aktual vs Prediksi Kota Probolinggo

Selain evaluasi visual, pengukuran kinerja model dilakukan secara kuantitatif menggunakan metrik *Mean Absolute Percentage Error* (MAPE) pada data uji. Pengujian dilakukan secara terpisah untuk setiap wilayah guna melihat konsistensi model dalam menangani karakteristik data yang berbeda-beda. Ringkasan hasil evaluasi untuk keseluruhan wilayah studi di Jawa Timur disajikan dalam Tabel 14.

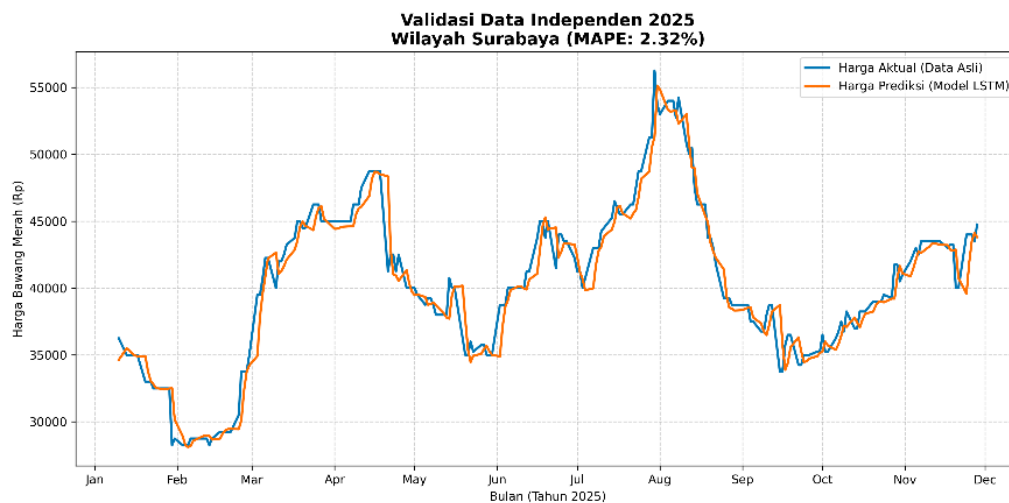
**Tabel. 14** Hasil pengujian Performa Antar Wilayah

| No | Nama Wilayah    | MAPE (%) | Kategori Akurasi |
|----|-----------------|----------|------------------|
| 1  | Kab. Banyuwangi | 4.1406%  | Sangat Baik      |
| 2  | Kota Surabaya   | 2.6390%  | Sangat Baik      |
| 3  | Kota Malang     | 4.6465%  | Sangat Baik      |
| 4  | Kota Kediri     | 2.2605%  | Sangat Baik      |

| No | Nama Wilayah     | MAPE (%) | Kategori Akurasi |
|----|------------------|----------|------------------|
| 5  | Kab. Jember      | 3.9588%  | Sangat Baik      |
| 6  | Kota Madiun      | 4.5632%  | Sangat Baik      |
| 7  | Kota Probolinggo | 2.4025%  | Sangat Baik      |
| 8  | Kab. Sumenep     | 2.3627%  | Sangat Baik      |
| 9  | Kota Blitar      | 3.4628%  | Sangat Baik      |

Berdasarkan Tabel 14, model LSTM menunjukkan performa yang sangat memuaskan di seluruh wilayah pengujian. Nilai error terendah dicatatkan oleh Kota Kediri dengan MAPE sebesar 2.2605%, sedangkan nilai error tertinggi berada pada Kota Malang sebesar 4.6465%. Meskipun terdapat variasi hasil, seluruh nilai MAPE berada jauh di bawah ambang batas 10%. Mengacu pada Tabel 3 Range Nilai MAPE, hasil didapat berada dalam kategori "Sangat Baik". Rata-rata MAPE untuk seluruh wilayah berada pada kisaran 3%, yang mengindikasikan bahwa model memiliki tingkat generalisasi yang tinggi.

Pengujian lanjutan dilakukan dengan menggunakan *independent dataset* yaitu data eksternal untuk memastikan bahwa model tidak *overfitting*. *Independent dataset* yang digunakan adalah data pada tahun 2025 yang tidak pernah digunakan untuk pelatihan ataupun pengujian, sehingga dapat menguji keandalan model pada data yang belum pernah dipelajari sebelumnya. Pengujian ini diterapkan pada masing-masing model di 9 wilayah penelitian. Hal ini bertujuan untuk memastikan bahwa kemampuan generalisasi model berlaku merata dan tidak hanya spesifik pada satu daerah saja. Sebagai representasi visual, Gambar 5 menampilkan hasil prediksi terhadap data aktual tahun 2025 untuk wilayah Kota Surabaya.



Gambar. 5 Grafik Validasi Model Independen Dataset Tahun 2025

Tabel. 15 Hasil Pengujian Independen Dataset Tahun 2025

| No | Nama Wilayah     | MAPE (%) | Kategori Akurasi |
|----|------------------|----------|------------------|
| 1  | Kab. Banyuwangi  | 3.7341%  | Sangat Baik      |
| 2  | Kota Surabaya    | 2.3243%  | Sangat Baik      |
| 3  | Kota Malang      | 4.0936%  | Sangat Baik      |
| 4  | Kota Kediri      | 2.1739%  | Sangat Baik      |
| 5  | Kab. Jember      | 2.2940%  | Sangat Baik      |
| 6  | Kota Madiun      | 5.2563%  | Sangat Baik      |
| 7  | Kota Probolinggo | 1.9398%  | Sangat Baik      |

| No | Nama Wilayah | MAPE (%) | Kategori Akurasi |
|----|--------------|----------|------------------|
| 8  | Kab. Sumenep | 2.3640%  | Sangat Baik      |
| 9  | Kota Blitar  | 3.6886%  | Sangat Baik      |
|    | Rata-rata    | 3.10%    | Sangat Baik      |

Untuk memberikan gambaran performa secara menyeluruh di seluruh lokasi penelitian, hasil evaluasi *error* (MAPE) pada dataset 2025 untuk 9 wilayah dirangkum dalam Tabel 15. Berdasarkan data pada Tabel 15, terlihat bahwa seluruh wilayah pengujian menghasilkan nilai MAPE yang tetap terjaga dalam kategori "Sangat Baik" (MAPE <10%) meskipun diuji menggunakan data tahun 2025 yang benar-benar baru. Secara keseluruhan, hasil penelitian ini sejalan dengan temuan pada penelitian terdahulu yang menyatakan bahwa algoritma *Deep Learning*, khususnya LSTM, bisa digunakan dalam memprediksi data runtun waktu yang fluktuatif [19]. Berdasarkan pengujian *independent dataset* menggunakan data tahun 2025, model LSTM yang diusulkan dalam penelitian ini menghasilkan nilai rata-rata MAPE 3.10%. Untuk memperjelas keandalan model dan kontribusi penelitian, hasil tersebut selanjutnya dibandingkan dengan penelitian terdahulu. Perbandingan dilakukan berdasarkan nilai evaluasi yang dilaporkan pada masing-masing penelitian terdahulu sesuai dengan dataset dan skenario pengujian yang digunakan oleh setiap peneliti. Hasil perbandingan dengan metode terdahulu dapat dilihat pada Tabel 16.

Berdasarkan Tabel 16, penelitian ini menghasilkan nilai MAPE sebesar 3.10% menggunakan metode LSTM dengan *hyperparameter tuning* pada sembilan wilayah di Jawa Timur, yang menunjukkan performa sangat baik dengan nilai MAPE di bawah 10%. Meskipun secara nilai MAPE penelitian ini berada di atas Sunariadi (2022) [4] dan Asnawi (2024) [7] yang masing-masing menggunakan metode SARIMA dan LSTM, perbandingan tersebut tidak sepenuhnya setara secara metodologis, mengingat kedua penelitian tersebut hanya menggunakan satu wilayah sebagai objek kajian. Hal serupa juga berlaku pada penelitian Adhwaningrum (2025) [20] dengan metode ARIMA yang memperoleh MAPE 5.72% serta Farisi (2025) [5] dengan Regresi Linier yang menghasilkan MAPE tertinggi sebesar 21.7%, yang keduanya pun hanya mencakup satu wilayah. Pengujian pada satu wilayah cenderung menghasilkan error yang lebih kecil karena homogenitas karakteristik data, sehingga model lebih mudah menyesuaikan dengan pola lokal dan berpotensi tidak mampu menggeneralisasi kondisi wilayah yang lebih beragam. Sebaliknya, penelitian ini membangun sembilan model independen yang masing-masing diuji menggunakan independent dataset, dan terbukti mampu menghasilkan nilai MAPE yang kompetitif pada berbagai wilayah, di mana nilai terendah diperoleh pada Kota Probolinggo sebesar 1.9398%, diikuti Kota Kediri sebesar 2.1739%, dan Kab. Jember sebesar 2.2940%, yang mengindikasikan bahwa model LSTM dengan *hyperparameter tuning* mampu menangkap pola data dengan sangat presisi. Dengan demikian, rata-rata MAPE sebesar 3.10% yang diperoleh mencerminkan konsistensi dan keandalan model secara menyeluruh di berbagai kondisi wilayah yang beragam, sehingga pendekatan yang diterapkan dalam penelitian ini tidak hanya kompetitif dari sisi akurasi, tetapi juga lebih unggul dalam hal cakupan wilayah.

**Tabel 16** Hasil Pengujian Independen Dataset Tahun 2025

| No | Peneliti         | Metode                              | Dataset                        | MAPE (%) | Keterangan                                       |
|----|------------------|-------------------------------------|--------------------------------|----------|--|
| 1  | Penelitian ini   | LSTM + <i>Hyperparameter tuning</i> | Sembilan wilayah di JAWA TIMUR | 3.10 %   | Menggunakan independent dataset dan multiwilayah |
| 2  | Farisi, 2025 [5] | Regresi Linier                      | Ponorogo                       | 21.7 %   | Menggunakan satu wilayah                         |

| No | Peneliti                | Metode | Dataset  | MAPE (%) | Keterangan               |
|----|-------------------------|--------|----------|----------|--------------------------|
| 3  | Adhwaningrum, 2025 [20] | ARIMA  | Brebes   | 5.72%    | Menggunakan satu wilayah |
| 4  | Sunariadi, 2022 [4]     | SARIMA | Nganjuk  | 2.01%    | Menggunakan satu wilayah |
| 5  | Asnawi, 2024 [7]        | LSTM   | Wonosobo | 1.80%    | Menggunakan satu wilayah |

#### 4. Kesimpulan dan Saran

Penelitian ini berhasil mengembangkan dan mengimplementasikan model *Long Short-Term Memory* (LSTM) untuk memprediksi harga harian bawang merah 9 daerah di wilayah Provinsi Jawa Timur, yaitu Kota Surabaya, Kota Malang, Kota Kediri, Kab. Jember, Kab. Banyuwangi, Kota Madiun, Kota Probolinggo, Kab. Sumenep, dan Kota Blitar. Berdasarkan hasil pengujian, ditemukan bahwa pola harga komoditas ini memiliki ketergantungan jangka pendek yang kuat dengan konfigurasi optimal pada *Window Size 7*, *Epoch 114*, *Batch Size 64*, dan *Dropout 0.02*. Secara keseluruhan, hasil pengujian pada *independent dataset* memberikan nilai MAPE dalam rentang 1.9398 % hingga 5.2563% dengan rata-rata MAPE pada sembilan wilayah sebesar 3.10%. Hasil ini termasuk dalam kategori "Sangat Baik". Meskipun demikian, penelitian ini memiliki keterbatasan pada penggunaan data time series harga yang belum mempertimbangkan faktor eksternal seperti pengaruh musim, curah hujan, atau hari libur serta jangkauan prediksi yang masih terbatas pada satu hari ke depan. Sebagai arah penelitian selanjutnya, perlu mengintegrasikan variabel eksternal seperti curah hujan, hari libur, volume panen, dan tingkat inflasi guna meningkatkan presisi model terhadap faktor musiman. Selain itu, pengembangan dapat diarahkan pada komparasi metode *Deep Learning* lain seperti GRU atau model *hybrid CNN-LSTM* untuk mencari arsitektur yang lebih efisien.

**Ucapan Terima Kasih:** Penulis menyampaikan terima kasih kepada Politeknik Negeri Malang atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini.

#### Referensi

- [1] Y. Ashari and A. Suhendar, "Implementasi Algoritma Long Short-Term Memory (Lstm) Untuk Memprediksi Harga Beras Di Jawa Tengah Berdasarkan Cuaca," *Djtechno: Jurnal Teknologi Informasi*, vol. 5, no. 3, pp. 624–636, 2024, <https://doi.org/10.46576/djtechno.v5i3.5136>.
- [2] Pusat Informasi Harga Pangan Strategis Nasional, "Tabel Harga Berdasarkan Komoditas Bawang Merah," <https://www.bi.go.id/hargapangan/TabelHarga/PasarTradisionalKomoditas>.
- [3] A. E. Radho, P. Sugiartawan, and G. A. Santiago, "Prediksi Jumlah Kasus COVID-19 Menggunakan Teknik Sliding Wondows dengan Metode BPNN," *Jurnal Sistem Informasi dan Komputer Terapan Indonesia (JSIKTI)*, vol. 4, no. 1, pp. 12–23, 2021, <https://doi.org/10.33173/jsikti.123>.
- [4] N. M. Sunariadi, P. K. Intan, D. C. R. Novitasari, and Y. Hariningsih, "Prediksi Produksi Bawang Merah Di Kabupaten Nganjuk Dengan Metode Seasonal Arima (Sarima)," *Transformasi : Jurnal Pendidikan Matematika dan Matematika*, vol. 6, no. 1, pp. 49–60, 2022, <https://doi.org/10.36526/tr.v6i1.1672>.
- [5] S. Al Farisi, G. A. Buntoro, and A. Prasetyo, "Prediksi Harga Bawang Merah dengan Regresi Linear Berbasis Website di Ponorogo," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 3, 2025, <https://doi.org/10.35957/jatisi.v12i3.13234>.
- [6] S. S. Dzikrina, M. Dihin, and K. Widya, "Perbandingan Metode Support Vector Machine dan Long Short Term Memory Dalam Prediksi Pendapatan Usaha Laundry," *Prosiding TAU SNARS-TEK Seminar Nasional Rekayasa dan Teknologi*, 2024. <https://doi.org/10.47970/snarstek.v2i1.709>
- [7] A. Triyadi, A. Suwondo, D. Asmarajati, N. Hasanah, and M. F. Asnawi, "Prediksi Harga Bawang Merah Kering Di Wonosobo Menggunakan Metode Long Short Term Memory," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 3, no. 2, pp. 133–138, 2024, <https://doi.org/10.55123/storage.v3i2.3601>.

- [8] M. Jahidul, I. Razin, M. A. Karim, M. F. Mridha, S. M. Rafiuddin, and T. Alam, "A Long Short-Term Memory (LSTM) Model for Business Sentiment Analysis Based on Recurrent Neural Network," Sep. 2025, Accessed: Feb. 20, 2026. <https://arxiv.org/pdf/2509.03060v1>
- [9] J. P. L. Caña, J. G. Herrero, and J. M. M. López, "Forecasting Nonlinear Systems with LSTM: Analysis and Comparison with EKF," *Sensors (Basel)*, vol. 21, no. 5, p. 1805, Mar. 2021, <https://doi.org/10.3390/s21051805>.
- [10] F. Kamalov and H. Sulieman, "Time series signal recovery methods: Comparative study," 2021 International Symposium on Networks, Computers and Communications, ISNCC 2021, 2021, <https://doi.org/10.1109/ISNCC52172.2021.9615669>.
- [11] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, <https://doi.org/10.14421/jiska.2024.9.3.178-191>.
- [12] M. Atharsyah, M. A. Romli, J. R. Utara, K. Sleman, and I. Yogyakarta, "Implementasi Model Lstm, Gru, Bilstm, Dan Bigru Dalam Prediksi Harga Nikel," *JURNAL RISET KOMPUTER (JURIKOM)*, vol. 12, no. 2, pp. 46–56, 2024. <https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/8495>
- [13] I. R. Nashrullah, E. R. Widasari, and B. H. Prasetyo, "Implementasi Metode Sliding Window Untuk Peningkatan Perekaman Real-Time Pada Sistem Deteksi Kesadaran Berbasis Sensor Electroencephalography Satu Kanal," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 3, pp. 1–11, 2026. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/16150>
- [14] R.; Abdul, P. A. Irma, and A. Irfan, "Penerapan Metode Long Short Term Memory (LSTM) Dalam Memprediksi Harga Saham Pt Bank Central Asia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 501–506, Feb. 2024. <https://doi.org/10.36040/jati.v8i1.8440>
- [15] T. Lattifia, P. Wira Buana, N. Kadek, and D. Rusjyanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM," *JITTER- Jurnal Ilmiah Teknologi dan Komputer*, vol. 3, no. 1, 2022. <http://ojs.unud.ac.id/index.php/jitter/article/view/85000>
- [16] W. M. Rahmat, N. Hidayat, and A. A. Soebroto, "Prediksi Penjualan Ponsel Pintar menggunakan Metode Jaringan Syaraf Tiruan Backpropagation Kombinasi Particle Swarm Optimization," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vo. 7, no. 1, 2023. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12213>
- [17] H. H. Nuha, "Mean Absolute Percentage Error (MAPE) dan Pengertiannya".
- [18] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *JOINS (Journal of Information System)*, vol. 5, no. 2, pp. 250–255, 2020, <https://doi.org/10.33633/joins.v5i2.3900>.
- [19] R. Muhammad, and I. Nurhaida, "Penerapan LSTM Dalam Deep Learning Untuk Prediksi Harga Kopi Jangka Pendek Dan Jangka Panjang," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 554–564, Jan. 2025, <https://doi.org/10.29100/jupi.v10i1.5904>.
- [20] A. S. Adhwaningrum, A. Ghufro, and Moch. F. Choirudin, "Implementasi Arima Dalam Peramalan Harga Bawang Merah Di Kabupaten Brebes," *Jurnal Statistika Industri dan Komputasi*, vol. 10, no. 02, pp. 34–42, Jul. 2025, <https://doi.org/10.34151/STATISTIKA.V10I02.5313>.