

Penerapan Model *Long Short-Term Memory* untuk Prediksi Harga Cabai Merah di Provinsi Banten

Rheno Budi Setiawan¹, Rendy Akbar Rianto², Muhamad Fikri Zulfikar³, Fandi Ali Mustika⁴,
Yuwan Jumaryadi^{5*}

^{1,2,3,4,5}Fakultas Ilmu Komputer, Universitas Mercu Buana, Indonesia

¹41820010086@student.mercubuana.ac.id, ²41820010015@student.mercubuana.ac.id,

³41820010077@student.mercubuana.ac.id, ⁴fandi.ali@mercubuana.ac.id,

^{5*}yuwan.jumaryadi@mercubuana.ac.id

Abstrak: Fluktuasi harga cabai merah memerlukan model prediksi yang mampu mendukung perencanaan produksi dan distribusi. Penelitian ini mengevaluasi LSTM univariat untuk prediksi satu hari ke depan harga cabai merah di Provinsi Banten menggunakan 730 data harian resmi periode Januari 2022–Desember 2023. Data diproses melalui forward fill, normalisasi MinMax berbasis data pelatihan, dan pembentukan sequence 30 hari. Sebanyak 700 sekuens dibagi secara kronologis menjadi 490 data pelatihan, 105 validasi, dan 105 pengujian. Model terdiri atas dua lapisan LSTM masing-masing 50 unit, dropout 0,2, Dense satu unit, optimizer Adam dengan learning rate 0,001, batch size 32, dan EarlyStopping. Pelatihan berhenti pada epoch ke-97 dengan validation loss terbaik 0,00028367 pada epoch ke-87. Hasil pengujian menghasilkan RMSE Rp3.498,29/kg, MAE Rp2.574,81/kg, dan MAPE 4,14%. Model mampu mengikuti arah umum perubahan harga, tetapi cenderung menghaluskan puncak dan mengalami keterlambatan ketika terjadi fluktuasi tajam. Kontribusi penelitian terletak pada evaluasi kontekstual data resmi Provinsi Banten dengan pembagian kronologis dan analisis periode kesalahan tinggi. Model dapat menjadi komponen awal sistem pemantauan harga, tetapi pengembangan selanjutnya perlu memasukkan variabel cuaca, produksi, distribusi, dan kebijakan serta membandingkannya dengan model baseline dan model deep learning lain.

Kata Kunci: Cabai Merah; Prediksi Harga; *Long Short-Term Memory*; Deret Waktu; Provinsi Banten.

Abstract: Red chili price volatility requires forecasting models that can support production and distribution planning. This study evaluates a univariate Long Short-Term Memory (LSTM) model for one-day-ahead forecasting of red chili prices in Banten Province using 730 official daily observations from January 2022 to December 2023. The data were processed using forward fill, training-based MinMax normalization, and 30-day sequence construction. The resulting 700 sequences were chronologically divided into 490 training, 105 validation, and 105 testing samples. The model comprised two LSTM layers with 50 units each, dropout layers of 0.2, one Dense output unit, the Adam optimizer with a learning rate of 0.001, a batch size of 32, and EarlyStopping. Training stopped at epoch 97, while the best validation loss of 0.00028367 was obtained at epoch 87. The model achieved an RMSE of IDR 3,498.29/kg, an MAE of IDR 2,574.81/kg, and

a MAPE of 4.14%. It captured the general price direction but tended to smooth peaks and lag during abrupt fluctuations. The study contributes a contextual evaluation using official Banten Province data, chronological splitting, and error analysis during highly volatile periods. The model may serve as an initial component of a price-monitoring system; however, future work should incorporate weather, production, distribution, and policy variables and compare the model against statistical and deep-learning baselines.

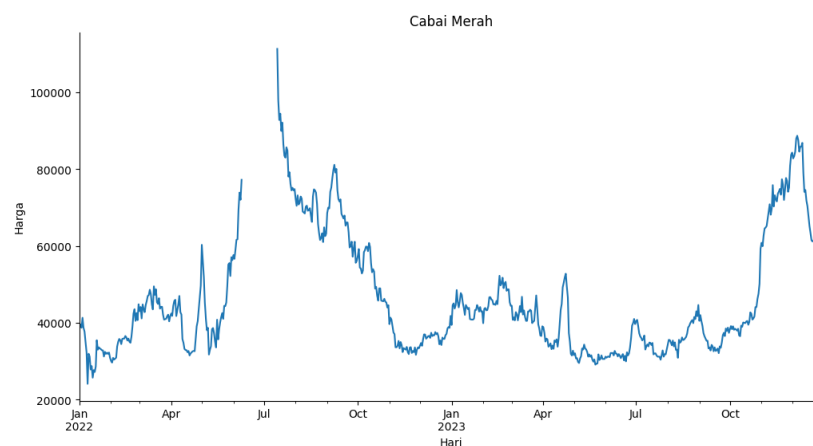
Keywords: Red Chili; Price Forecasting; *Long Short-Term Memory*; Time Series; Banten Province

1. PENDAHULUAN

Cabai merah (*Capsicum annum* L.) merupakan salah satu komoditas hortikultura semusim yang memiliki nilai ekonomi penting di Indonesia. Selain memenuhi kebutuhan konsumsi domestik, komoditas ini juga berperan dalam perdagangan dan memiliki potensi ekspor. Di Provinsi Banten, cabai merah termasuk komoditas hortikultura unggulan dengan pergerakan harga yang cenderung berfluktuasi. Perubahan harga tersebut dipengaruhi oleh berbagai faktor, antara lain kondisi cuaca, keseimbangan penawaran dan permintaan, kelancaran distribusi, serta kebijakan pemerintah [1].

Fluktuasi harga cabai merah selama periode Januari 2022 hingga Desember 2023 ditunjukkan pada Gambar 1. Harga mengalami perubahan yang cukup tajam, mulai dari kisaran Rp25.000 per kilogram hingga lebih dari Rp100.000 per kilogram. Kenaikan tertinggi terlihat pada pertengahan 2022, kemudian diikuti penurunan secara bertahap hingga akhir tahun. Pada 2023, harga relatif lebih rendah pada beberapa periode, tetapi kembali mengalami peningkatan menjelang akhir tahun. Pola tersebut menunjukkan bahwa harga cabai merah bersifat dinamis dan dapat mengalami perubahan dalam waktu relatif singkat.

Salah satu faktor yang dapat memengaruhi perubahan produksi dan pasokan cabai adalah variabilitas musim. Indonesia memiliki musim hujan dan musim kemarau, tetapi waktu awal, durasi, dan intensitasnya dapat berubah dari tahun ke tahun akibat pengaruh sistem monsun, kondisi lokal, serta dinamika atmosfer-laut berskala global, seperti El Niño, La Niña, dan *Indian Ocean Dipole*. Variabilitas tersebut dapat memengaruhi produktivitas, distribusi, dan ketersediaan cabai di pasar sehingga turut berkontribusi terhadap fluktuasi harga.



Gambar 1. Perkembangan Harga Cabai Merah di Provinsi Banten Periode Januari 2022–Desember 2023

Tingginya fluktuasi harga sebagaimana ditunjukkan pada Gambar 1 menegaskan perlunya model yang mampu mempelajari pola temporal dan hubungan nonlinier pada data harga. LSTM telah banyak digunakan dalam peramalan deret waktu karena mampu mempertahankan informasi penting dari periode sebelumnya [2], [3], [4], [5], [6]. Namun, kajian mutakhir menunjukkan bahwa penelitian prediksi harga komoditas terus berkembang dari model tunggal menuju model hibrida, dekomposisi sinyal, mekanisme perhatian, dan integrasi variabel eksternal untuk meningkatkan ketahanan model terhadap volatilitas. Penelitian mengenai harga cabai di Indonesia juga telah dilakukan, tetapi sebagian besar menggunakan wilayah, periode, konfigurasi, dan skema evaluasi yang berbeda.

Posisi penelitian ini berada pada evaluasi model LSTM univariat berbasis data harga harian resmi pada konteks Provinsi Banten. Berbeda dari penelitian terdahulu yang menggunakan data Kota Malang dengan konfigurasi sequence 21 hari dan proporsi pelatihan-pengujian 70:30 [7], penelitian ini menggunakan 730 observasi harian periode 2022–2023, sequence 30 hari, pembagian kronologis pelatihan-validasi-pengujian, serta EarlyStopping untuk menjaga generalisasi model. Kontribusi utama penelitian terletak pada: (1) penyediaan bukti empiris spesifik Provinsi Banten menggunakan sumber data pemerintah; (2) penerapan alur pemodelan yang mencegah kebocoran data melalui pembagian kronologis dan normalisasi berbasis data pelatihan; serta (3) analisis kesalahan model pada periode perubahan harga tajam, bukan hanya pelaporan metrik agregat. Kontribusi tersebut bersifat kontekstual dan eksperimental, bukan pengusulan arsitektur LSTM baru. LSTM telah banyak diterapkan pada berbagai permasalahan prediksi deret waktu, termasuk prediksi harga cabai [7], [8], harga saham [9], [10], dan harga komoditas pangan [11].

Penelitian ini bertujuan mengevaluasi kemampuan model LSTM dalam memprediksi harga cabai merah harian di Provinsi Banten berdasarkan data Panel Harga Pangan Badan Pangan Nasional periode Januari 2022 hingga Desember 2023. Kinerja model dievaluasi menggunakan RMSE, MAE, dan MAPE, serta dianalisis melalui kurva pelatihan dan perbandingan harga aktual-prediksi. Hasil penelitian diharapkan memberikan dasar empiris bagi pengembangan sistem peringatan dini harga dan mendukung perencanaan produksi, distribusi, serta perdagangan cabai merah

2. METODE PENELITIAN

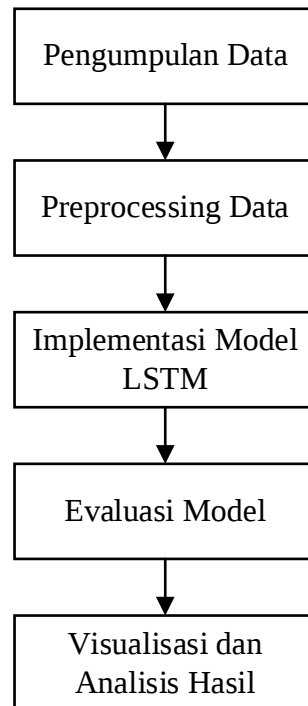
2.1. Sumber Data

Penelitian ini menggunakan data sekunder berupa harga eceran cabai yang diperoleh dari Panel Harga Pangan Badan Pangan Nasional. Data tersedia pada menu Tabel Harga dengan memilih kategori Harga Eceran. Informasi yang digunakan dalam penelitian meliputi tanggal pencatatan, wilayah atau provinsi, jenis komoditas, dan harga komoditas dalam satuan rupiah per kilogram. Penelitian difokuskan pada data harga cabai di Provinsi Banten selama periode 1 Januari 2022 sampai dengan 31 Desember 2023.

2.2. Teknik Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik *web scraping* pada laman panelharga.badanpangan.go.id. Teknik tersebut digunakan untuk mengekstraksi data harga pangan secara otomatis dari halaman web dan menyusunnya ke dalam format tabular. Data yang dikumpulkan meliputi tanggal pencatatan, nama komoditas, provinsi, dan harga eceran. Hasil pengumpulan data selanjutnya disimpan untuk digunakan pada tahap pembersihan, transformasi, pemodelan, dan evaluasi.

2.3. Tahapan Penelitian



Gambar 2. Tahapan Penelitian

Berdasarkan Gambar 2, tahapan penelitian terdiri atas pengumpulan data, prapemrosesan data, implementasi model LSTM, evaluasi model, dan visualisasi data. Penjelasan setiap tahapan tersebut disajikan sebagai berikut.

2.3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data historis harga cabai yang diperoleh dari Panel Harga Pangan Badan Pangan Nasional melalui teknik *web scraping*. Data yang dikumpulkan mencakup tanggal pencatatan, provinsi, jenis komoditas, dan harga eceran dalam satuan rupiah per kilogram. Penelitian difokuskan pada data harga cabai di Provinsi Banten selama periode pengamatan yang telah ditentukan.

Data hasil *web scraping* selanjutnya disusun dalam format tabular dan diurutkan berdasarkan tanggal pencatatan. Penyusunan data secara kronologis dilakukan karena model *Long Short-Term Memory* (LSTM) membutuhkan data berurutan untuk mempelajari pola perubahan harga dari waktu ke waktu.

2.3.2. Preprocessing Data

Tahap *preprocessing* meliputi konversi tanggal, pengurutan kronologis, penanganan nilai hilang, normalisasi, pembentukan sekuens, dan pembagian data. Nilai hilang pada harga cabai merah diisi menggunakan *forward fill* karena data tersusun harian dan nilai terdekat sebelumnya dipertahankan sebagai pendekatan imputasi yang sederhana. Pemodelan dilakukan secara univariat menggunakan harga cabai merah sebagai fitur historis dan target prediksi.

Setelah 34 nilai harga yang hilang ditangani, diperoleh 730 observasi harian lengkap. Data kemudian dibentuk menjadi sekuens sepanjang 30 hari. Pemilihan sequence 30 hari dimaksudkan untuk merepresentasikan pola jangka pendek sekitar satu bulan dan

menyediakan konteks temporal yang cukup tanpa mengurangi jumlah sampel secara berlebihan. Dari 730 observasi dihasilkan 700 pasangan input–target, karena 30 observasi pertama digunakan sebagai jendela awal prediksi.

Normalisasi dilakukan menggunakan MinMaxScaler pada rentang 0–1 sebagaimana dirumuskan pada Persamaan (1). Scaler hanya di-*fit* pada data pelatihan, kemudian parameter minimum dan maksimum yang sama digunakan untuk mentransformasikan data validasi dan pengujian. Prosedur ini diterapkan untuk mencegah *data leakage* dari periode masa depan ke proses pelatihan.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Pada persamaan (1), z_i sebagai nilai setelah standardisasi, x_i sebagai nilai asli, μ sebagai nilai rata-rata data pelatihan, dan σ sebagai standar deviasi data pelatihan. Untuk mencegah kebocoran data, nilai rata-rata dan standar deviasi dihitung hanya berdasarkan data pelatihan. Parameter yang sama kemudian digunakan untuk mentransformasikan data pengujian.

Data sekuens dibagi secara kronologis tanpa pengacakan menjadi 490 sampel pelatihan (70%), 105 sampel validasi (15%), dan 105 sampel pengujian (15%). Proporsi 70% dipilih agar model memperoleh data yang memadai untuk mempelajari pola, sedangkan masing-masing 15% dialokasikan untuk pemilihan model dan evaluasi akhir yang independen. Bentuk input akhir adalah (490, 30, 1), (105, 30, 1), dan (105, 30, 1).

2.3.3. Implementasi Model LSTM

Sebagaimana dirangkum pada Tabel 1, model dibangun menggunakan API *Sequential* pada Keras dengan ukuran masukan (30, 1), yang merepresentasikan 30 hari pengamatan dan satu fitur harga. Arsitektur model terdiri atas dua lapisan LSTM yang masing-masing memiliki 50 unit, dua lapisan *dropout* dengan nilai 0,2, serta satu lapisan *Dense* dengan satu unit keluaran. Lapisan LSTM pertama menggunakan *return_sequences=True* agar seluruh keluaran pada setiap langkah waktu dapat diteruskan ke lapisan LSTM kedua. Sementara itu, lapisan LSTM kedua menggunakan *return_sequences=False* untuk menghasilkan representasi akhir dari sekuens masukan.

Jumlah 50 unit dipilih sebagai konfigurasi menengah untuk menjaga kapasitas model tetap memadai pada dataset yang relatif terbatas. Dropout 0,2 digunakan sebagai regularisasi untuk mengurangi risiko overfitting. Model dikompilasi menggunakan Adam dengan learning rate 0,001 dan fungsi loss mean squared error. Adam dipilih karena pembaruan learning rate adaptifnya sesuai untuk optimasi jaringan berulang. Pelatihan menggunakan batch size 32 dan maksimum 200 epoch.

EarlyStopping diterapkan dengan *monitor=val_loss*, *patience=10*, dan *restore_best_weights=True*. Mekanisme ini menghentikan pelatihan ketika validation loss tidak membaik selama 10 epoch berturut-turut dan mengembalikan bobot pada epoch terbaik. Penggunaan data validasi dan EarlyStopping dimaksudkan untuk membatasi overfitting serta memastikan bahwa model akhir dipilih berdasarkan performa pada data yang tidak digunakan untuk pembaruan bobot.

Tabel 1. Konfigurasi dan Pertimbangan Parameter Model LSTM

Parameter	Nilai	Pertimbangan
Sequence length	30 hari	Merepresentasikan pola jangka pendek sekitar satu bulan
LSTM units	50 + 50	Kapasitas menengah untuk dataset 730 observasi

Dropout	0,2	Regularisasi untuk mengurangi overfitting
Batch size	32	Kompromi stabilitas gradien dan efisiensi komputasi
Optimizer	Adam	Optimasi adaptif untuk jaringan berulang
Learning rate	0,001	Nilai awal umum yang stabil untuk Adam
Maksimum epoch	200	Memberi ruang konvergensi dengan kontrol EarlyStopping
EarlyStopping	patience=10	Menghentikan pelatihan saat validasi tidak membaik

2.3.4. Evaluasi Model

Evaluasi dilakukan pada 105 sampel pengujian setelah hasil prediksi dikembalikan ke skala rupiah melalui inverse transform. Metrik yang digunakan adalah Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE). Ketiga metrik digunakan secara bersama karena RMSE lebih sensitif terhadap kesalahan besar, MAE menunjukkan rata-rata selisih dalam satuan rupiah, sedangkan MAPE menunjukkan kesalahan relatif dalam persentase.

RMSE dihitung sebagai akar dari rata-rata kuadrat selisih antara nilai aktual dan prediksi [12]. Nilai RMSE yang lebih rendah menunjukkan prediksi yang lebih dekat dengan harga aktual, tetapi metrik ini memberikan penalti lebih besar pada kesalahan ekstrem [13]. MAE mengukur rata-rata selisih absolut dalam satuan rupiah, sedangkan MAPE mengukur rata-rata kesalahan absolut relatif terhadap nilai actual [3]. Nilai MAPE yang lebih rendah menunjukkan kesalahan proporsional yang lebih kecil. MAPE dihitung menggunakan persamaan (2).

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

Pada kedua persamaan tersebut, y_i merupakan nilai aktual, $\hat{y}_i$ merupakan nilai hasil prediksi, dan n merupakan jumlah data pengujian. Semakin rendah nilai MSE dan MAPE, semakin baik kemampuan model dalam memprediksi harga cabai.

2.3.5. Visualisasi Data

Hasil prediksi divisualisasikan menggunakan pustaka Matplotlib dalam bentuk grafik garis. Grafik menampilkan perbandingan antara data harga aktual, hasil prediksi pada data pelatihan, dan hasil prediksi pada data pengujian. Visualisasi tersebut digunakan untuk mengamati kemampuan model dalam mengikuti pola perubahan harga serta mengetahui perbedaan antara nilai aktual dan nilai prediksi.

Selain dalam bentuk grafik, hasil prediksi disajikan dalam bentuk tabel yang memuat tanggal, harga aktual, harga hasil prediksi, dan nilai kesalahan prediksi. Penyajian grafik dan tabel digunakan untuk mendukung interpretasi hasil evaluasi serta memberikan gambaran yang lebih rinci mengenai kinerja model LSTM dalam memprediksi harga cabai

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data penelitian diperoleh dari situs resmi pemerintah, yaitu Panel Harga Pangan Badan Pangan Nasional dan Data Online Badan Meteorologi, Klimatologi, dan Geofisika. Dataset mencakup periode 1 Januari 2022 hingga 31 Desember 2023 dengan total 730 observasi

harian. Variabel yang tersedia meliputi tanggal, cabai_merah, curah_hujan, dan lamanya_penyinaran_matahari.

Sebagaimana disajikan pada Tabel 2, kolom tanggal menunjukkan waktu pencatatan data, kolom cabai_merah menunjukkan harga cabai merah dalam rupiah per kilogram, sedangkan kolom curah_hujan dan lamanya_penyinaran_matahari menunjukkan kondisi cuaca pada periode yang sama. Beberapa observasi masih mengandung nilai yang hilang (NaN), khususnya pada variabel cuaca, sehingga diperlukan tahap prapemrosesan sebelum pemodelan dilakukan. Dalam penelitian ini, variabel cabai_merah digunakan sebagai fitur utama sekaligus target prediksi model LSTM.

Tabel 2. Dataset

	tanggal	cabai_merah	curah_hujan	lamanya_penyinaran_matahari
0	1/1/2022	39240.0	0.7	5.1
1	2/1/2022	39570.0	0.0	8.5
2	3/1/2022	38700.0	NaN	3.2
3	4/1/2022	41310.0	NaN	9.1
...	...	...	...	...
729	31/12/2023	61260.0	7.2	1.9

3.2. Preprocessing Data

Tahap prapemrosesan dilakukan untuk memastikan kualitas, konsistensi, dan kesesuaian struktur data dengan kebutuhan model LSTM. Proses ini mencakup konversi format tanggal, penanganan nilai yang hilang, normalisasi, pembentukan sekuens, serta pembagian data menjadi data pelatihan, validasi, dan pengujian.

3.2.1. Konversi Tanggal dan Penanganan Missing Values

Kolom tanggal dikonversi ke tipe datetime menggunakan parameter `dayfirst=True` karena format tanggal pada dataset mengikuti susunan hari/bulan/tahun. Konversi tersebut diperlukan agar data dapat diurutkan secara kronologis dan diproses sebagai deret waktu.

Nilai yang hilang pada kolom cabai_merah ditangani menggunakan metode *forward fill* (ffill), yaitu mengisi nilai kosong dengan nilai valid terakhir berdasarkan urutan waktu. Metode ini dipilih karena harga pada suatu periode memiliki keterkaitan temporal dengan periode sebelumnya. Sebagaimana ditunjukkan pada Tabel 3, terdapat 34 nilai yang hilang pada kolom cabai_merah sebelum prapemrosesan. Setelah penerapan *forward fill*, seluruh nilai yang hilang pada kolom tersebut berhasil diisi.

Nilai yang hilang pada kolom curah_hujan dan lamanya_penyinaran_matahari tidak ditangani lebih lanjut karena kedua variabel tersebut tidak digunakan sebagai input model. Dengan demikian, pemodelan dilakukan secara univariat dengan menggunakan cabai_merah sebagai fitur historis sekaligus variabel target.

Tabel 3. Missing Values

	Sebelum	Sesudah
tanggal	0	0
cabai_merah	34	0
curah_hujan	331	331
lamanya_penyinaran_matahari	1	1

3.2.2. Normalisasi Data

Normalisasi diterapkan pada kolom cabai_merah menggunakan MinMaxScaler untuk mentransformasikan nilai harga ke dalam rentang 0 hingga 1. Proses ini dilakukan agar skala data lebih seragam dan proses optimasi model LSTM berlangsung lebih stabil. Untuk mencegah *data leakage*, parameter normalisasi sebaiknya dihitung hanya berdasarkan data pelatihan, kemudian digunakan untuk mentransformasikan data validasi dan pengujian.

Contoh hasil transformasi empat observasi awal disajikan pada Tabel 4. Nilai 0,173544, 0,177327, 0,167354, dan 0,197272 merupakan hasil normalisasi harga cabai merah ke dalam skala 0–1.

Tabel 4. Hasil Normalisasi

	cabai_merah
0	0.173544
1	0.177327
2	0.167354
3	0.197272

3.2.3. Pembentukan Sekuens Data

Data yang telah dinormalisasi selanjutnya diubah menjadi pasangan input X dan target y menggunakan fungsi `create_sequences`. Panjang sekuens ditetapkan sebesar 30 hari. Dengan konfigurasi tersebut, setiap sampel input terdiri atas harga cabai merah selama 30 hari sebelumnya, sedangkan targetnya adalah harga pada hari berikutnya.

Sebagai contoh, harga pada hari ke-1 hingga hari ke-30 digunakan untuk memprediksi harga pada hari ke-31. Selanjutnya, harga pada hari ke-2 hingga hari ke-31 digunakan untuk memprediksi harga pada hari ke-32. Proses tersebut dilakukan secara bergeser hingga seluruh data selesai dibentuk menjadi sekuens.

Karena model menggunakan satu variabel input, data kemudian disusun dalam format tiga dimensi yang dibutuhkan oleh LSTM, yaitu:

(jumlah sampel, langkah waktu, jumlah fitur)

Dengan demikian, setiap sampel memiliki 30 langkah waktu dan satu fitur berupa harga cabai merah.

3.2.4. Pembagian Data

Sebagaimana pada Tabel 5, sebanyak 700 sekuens dibagi secara kronologis menjadi 490 sampel pelatihan (70%), 105 sampel validasi (15%), dan 105 sampel pengujian (15%). Pembagian dilakukan tanpa pengacakan untuk mempertahankan urutan temporal dan mencegah masuknya informasi dari periode masa depan ke dalam data pelatihan.

Data pelatihan digunakan untuk pembaruan bobot, data validasi untuk memantau `val_loss` dan menentukan epoch terbaik, sedangkan data pengujian hanya digunakan pada evaluasi akhir. Bentuk X_{train} , X_{val} , dan X_{test} masing-masing adalah (490, 30, 1), (105, 30, 1), dan (105, 30, 1).

Tabel 5. Pembagian Dataset Setelah Pembentukan Sekuens

Subset	Proporsi	Jumlah sekuens	Bentuk input
Pelatihan	70%	490	(490, 30, 1)
Validasi	15%	105	(105, 30, 1)
Pengujian	15%	105	(105, 30, 1)
Total	100%	700	(700, 30, 1)

3.3. Implementasi Model LSTM

Hasil pelatihan model LSTM disajikan pada Tabel 6, yang menunjukkan perkembangan nilai *training loss* dan *validation loss* pada beberapa *epoch* representatif. Model memiliki total 30.651 parameter yang seluruhnya termasuk parameter yang dapat dilatih. Proses pelatihan ditetapkan maksimum sebanyak 200 *epoch*, tetapi dihentikan pada *epoch* ke-97 melalui mekanisme *EarlyStopping*.

Tabel 6. Perkembangan Training Loss dan Validation Loss Model LSTM

Epoch	Training Loss	Validation Loss	Keterangan
1	0,0370	0,002000	Awal pelatihan
2	0,0102	0,004200	Terjadi fluktuasi awal
3	0,0069	0,001200	Loss mulai menurun
...	...	...	...
...	...	...	...
85	0,0017	0,000285	Mendekati nilai terbaik
87	0,0019	0,000284	Validation loss terbaik
90	0,0018	0,000329	Tidak terjadi perbaikan

Pada awal pelatihan, nilai *training loss* sebesar 0,0370 dan *validation loss* sebesar 0,002000. Nilai *training loss* kemudian menurun secara bertahap hingga mencapai sekitar 0,0017 pada akhir pelatihan. Sementara itu, nilai *validation loss* terbaik sebesar 0,00028367 diperoleh pada *epoch* ke-87. Setelah *epoch* tersebut, tidak terjadi perbaikan nilai *validation loss* selama 10 *epoch* berturut-turut, sehingga proses pelatihan dihentikan dan bobot model dikembalikan ke kondisi terbaik pada *epoch* ke-87.

Pola penurunan *training loss* dan *validation loss* pada Tabel 6 menunjukkan bahwa model mampu mempelajari pola temporal pada data harga cabai merah dengan baik. Meskipun terdapat fluktuasi pada beberapa *epoch*, nilai *validation loss* tetap berada pada tingkat yang rendah dan tidak menunjukkan peningkatan secara terus-menerus. Kondisi tersebut menunjukkan bahwa model tidak memperlihatkan indikasi *overfitting* yang kuat selama proses pelatihan.

3.4. Evaluasi Model

Sebagaimana disajikan pada Tabel 7, hasil pengujian model menghasilkan nilai RMSE sebesar Rp3.498,29/kg, MAE sebesar Rp2.574,81/kg, dan MAPE sebesar 4,14%. Nilai MAE menunjukkan bahwa rata-rata selisih absolut antara harga hasil prediksi dan harga aktual adalah sekitar Rp2.574,81/kg. Sementara itu, nilai RMSE yang lebih tinggi daripada MAE menunjukkan bahwa masih terdapat beberapa kesalahan prediksi yang relatif besar pada periode tertentu. Kondisi tersebut mengindikasikan bahwa distribusi kesalahan prediksi belum sepenuhnya seragam.

Nilai MAPE sebesar 4,14% menunjukkan tingkat kesalahan relatif yang rendah pada periode pengujian. Kinerja tersebut kemungkinan didukung oleh adanya autokorelasi pada harga harian, penggunaan sekuens data selama 30 hari sebagai masukan untuk memprediksi harga pada hari berikutnya, proses normalisasi, serta pembagian data secara kronologis. Namun, nilai MAPE tidak hanya mencerminkan kemampuan arsitektur LSTM. Karakteristik periode pengujian, tingkat volatilitas harga, horizon prediksi satu langkah ke depan, dan rentang harga juga memengaruhi besarnya nilai MAPE. Oleh karena itu, perbandingan hasil antarpelelitian perlu dilakukan secara hati-hati.

Dibandingkan penelitian harga cabai di Kota Malang yang melaporkan MSE 0,0155 pada data ternormalisasi [7], hasil penelitian ini tidak dapat dibandingkan langsung karena

perbedaan skala, wilayah, sequence, dan pembagian data. Sebagai konteks internasional, stacked LSTM untuk harga jintan harian menghasilkan MAPE 12,25% pada horizon satu tahun [14], sedangkan model hibrida VMD-EEMD-LSTM dilaporkan meningkatkan akurasi pada beberapa komoditas pertanian mingguan [15]. Perbedaan tersebut menegaskan bahwa performa model dipengaruhi oleh karakteristik komoditas, horizon prediksi, volatilitas, serta kompleksitas fitur. Dengan konfigurasi univariat dan horizon satu hari, MAPE 4,14% menunjukkan hasil yang kompetitif, tetapi belum membuktikan keunggulan terhadap model lain karena penelitian ini belum melakukan uji baseline yang setara.

Tabel 7. Hasil Evaluasi Kinerja Model LSTM

Metrik	Nilai	Interpretasi
RMSE	3.498,29	Kesalahan lebih besar masih muncul pada beberapa periode
MAE	2.574,81	Rata-rata selisih prediksi sekitar Rp2.574,81/kg
MAPE	4,14%	Kinerja prediksi sangat baik

3.5. Visualisasi Data

Visualisasi dilakukan untuk memperkuat interpretasi terhadap kinerja model LSTM. Hasil visualisasi disajikan dalam bentuk grafik perbandingan harga aktual dan harga hasil prediksi serta kurva *loss* pada data pelatihan dan validasi.

Gambar 3 menunjukkan perbandingan antara harga cabai merah aktual pada data pengujian dan harga yang diprediksi oleh model LSTM. Secara umum, garis prediksi mampu mengikuti arah dan pola perubahan harga aktual, baik pada periode kenaikan maupun penurunan harga. Pada awal periode pengujian, harga berada pada kisaran Rp32.000–Rp40.000 per kilogram dan hasil prediksi relatif berdekatan dengan nilai aktual. Ketika harga mengalami kenaikan tajam pada sekitar hari ke-40 hingga mencapai lebih dari Rp60.000 per kilogram, model masih dapat mengikuti perubahan tersebut, meskipun hasil prediksi terlihat lebih halus dan sedikit tertinggal dibandingkan nilai aktual.

Kesalahan prediksi terlihat lebih besar ketika harga mengalami kenaikan tajam, khususnya sekitar hari ke-40 hingga ke-85 pada data pengujian. Pada periode tersebut, garis prediksi cenderung lebih halus dan berada di bawah beberapa puncak aktual. Pola ini menunjukkan kecenderungan model menghasilkan lag dan meredam amplitudo perubahan ekstrem karena prediksi terutama dibentuk dari pola historis 30 hari. Lonjakan harga yang dipengaruhi kejadian eksternal tidak tersedia sebagai fitur, sehingga model tidak memperoleh sinyal langsung mengenai perubahan pasokan, musim panen, distribusi, atau intervensi kebijakan.

Pada periode harga yang bergerak lebih stabil, jarak antara garis aktual dan prediksi relatif kecil, sedangkan selisih meningkat saat perubahan harga berlangsung cepat. Temuan visual ini konsisten dengan selisih RMSE dan MAE. Untuk memperkuat analisis, penelitian lanjutan perlu menyajikan tabel lima atau sepuluh tanggal dengan absolute error tertinggi beserta konteks peristiwa pasar pada tanggal tersebut.

Gambar 4 memperlihatkan perubahan nilai *loss* pada data pelatihan dan validasi selama proses pelatihan model. Pada awal pelatihan, nilai *training loss* mengalami penurunan yang sangat tajam dari sekitar 0,037 menjadi kurang dari 0,010. Setelah beberapa *epoch*, penurunan berlangsung lebih lambat dan cenderung stabil hingga akhir pelatihan. Pola tersebut menunjukkan bahwa model dapat mempelajari pola dasar data dengan cepat pada tahap awal, kemudian melakukan penyesuaian bobot secara bertahap pada *epoch* berikutnya.

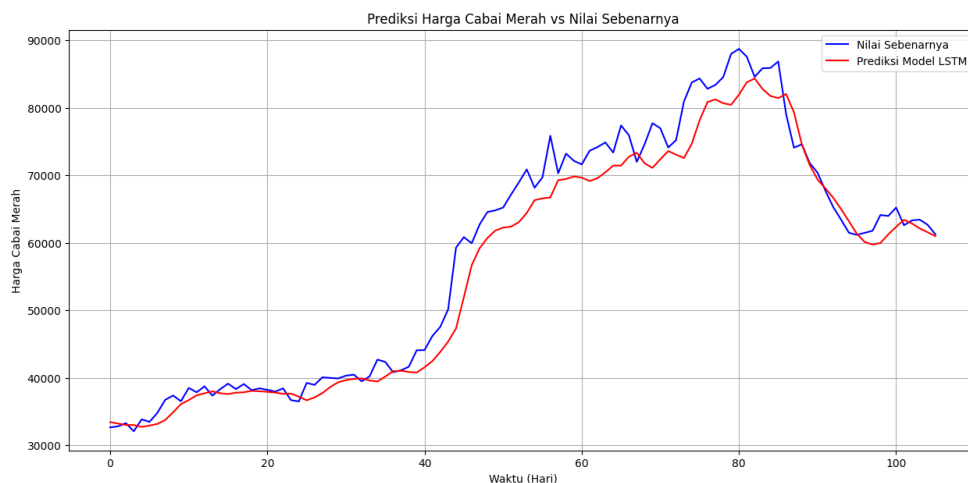
Nilai *validation loss* juga menurun dan bertahan pada tingkat yang rendah tanpa menunjukkan peningkatan secara berkelanjutan. Ketika training loss terus menurun dan validation loss tetap stabil, model tidak menunjukkan indikasi overfitting yang kuat. Pola

tersebut mengindikasikan bahwa model masih mampu mempertahankan kemampuan generalisasi terhadap data yang tidak digunakan dalam pembaruan bobot.

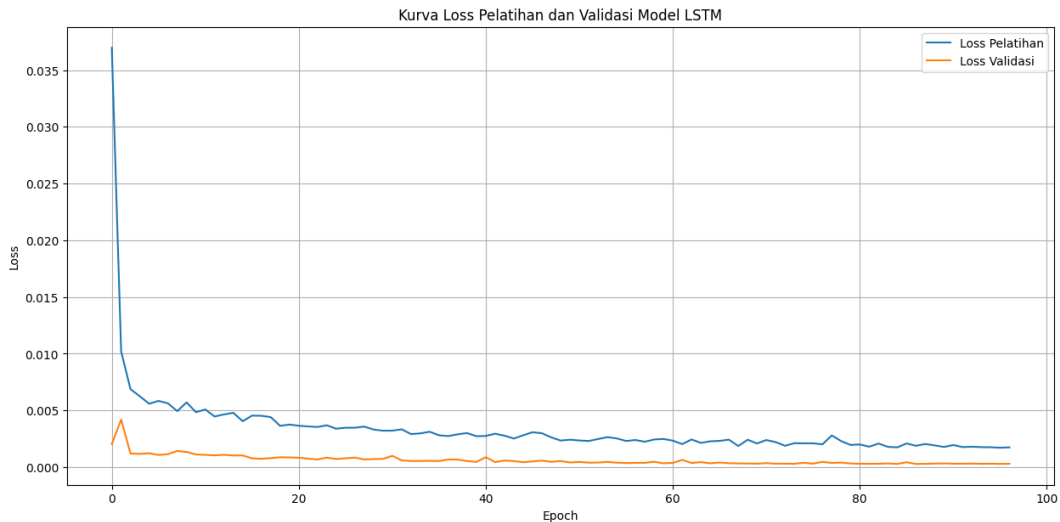
Nilai validation loss yang lebih rendah daripada training loss kemungkinan dipengaruhi oleh penggunaan lapisan dropout. Selama proses pelatihan, dropout menonaktifkan sejumlah unit secara acak sehingga proses pembelajaran menjadi lebih sulit dan nilai training loss dapat menjadi lebih tinggi. Sebaliknya, pada tahap validasi, seluruh unit jaringan digunakan tanpa menonaktifkan dropout, sehingga model dapat menghasilkan nilai validation loss yang lebih rendah. Temuan ini sejalan dengan hasil penelitian Khan et al. [16], yang menunjukkan bahwa penerapan *dropout* sebagai teknik regularisasi dapat mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model.

Selain itu, Santra et al. [15] menunjukkan bahwa penerapan *dropout* dapat meningkatkan kemampuan generalisasi model dengan mengurangi ketergantungan yang berlebihan antarneuron. Dengan demikian, kestabilan validation loss dalam penelitian ini tidak hanya menunjukkan proses pelatihan yang relatif baik, tetapi juga mengindikasikan bahwa kombinasi lapisan dropout dan mekanisme EarlyStopping mampu membantu membatasi risiko *overfitting*.

Berdasarkan pola kedua kurva tersebut, proses pelatihan dapat dinilai berlangsung stabil dan model memiliki kemampuan generalisasi yang baik terhadap data validasi. Kurva *loss* yang cenderung mendatar pada *epoch* akhir juga menunjukkan bahwa peningkatan jumlah *epoch* kemungkinan hanya memberikan perbaikan yang terbatas.



Gambar 3. Perbandingan harga cabai merah aktual dan hasil prediksi model LSTM pada data pengujian



Gambar 4. Kurva *loss* pelatihan dan validasi model LSTM

3.6. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, model hanya menggunakan harga historis sebagai variabel input sehingga belum mempertimbangkan faktor eksternal, seperti curah hujan, musim panen, volume produksi, kelancaran distribusi, inflasi, hari besar, dan kebijakan stabilisasi harga. Kedua, periode data hanya mencakup dua tahun sehingga variasi musiman dan kejadian ekstrem yang dipelajari model masih terbatas. Ketiga, evaluasi menggunakan satu pembagian kronologis dan belum menerapkan *walk-forward validation*. Keempat, penelitian belum membandingkan LSTM dengan baseline seperti *naive forecast*, ARIMA, GRU, atau model hibrida. Oleh karena itu, nilai MAPE yang rendah perlu ditafsirkan sebagai performa pada konfigurasi dan periode pengujian ini, bukan sebagai bukti bahwa LSTM selalu lebih unggul daripada metode lain.

4. KESIMPULAN

Penelitian ini memberikan bukti empiris bahwa LSTM univariat dapat digunakan untuk prediksi satu hari ke depan harga cabai merah di Provinsi Banten menggunakan data harian resmi periode 2022–2023. Kontribusi penelitian terletak pada penerapan alur pemodelan kronologis yang memisahkan 490 sampel pelatihan, 105 validasi, dan 105 pengujian, normalisasi berbasis data pelatihan, serta analisis kesalahan pada periode perubahan harga tajam. Model mencapai RMSE Rp3.498,29/kg, MAE Rp2.574,81/kg, dan MAPE 4,14%, serta mampu mengikuti arah umum pergerakan harga. Namun, model cenderung menghaluskan puncak dan mengalami keterlambatan ketika terjadi lonjakan mendadak karena hanya menggunakan informasi harga historis. Secara praktis, hasil penelitian ini dapat menjadi komponen awal dalam pengembangan sistem pemantauan atau peringatan harga. Namun, hasil tersebut belum dapat digunakan sebagai satu-satunya dasar dalam pengambilan kebijakan. Penelitian selanjutnya perlu mengintegrasikan variabel cuaca, produksi, musim panen, distribusi, inflasi, hari besar, dan kebijakan pemerintah. Selain itu, penelitian selanjutnya dapat menerapkan metode *walk-forward validation* serta membandingkan kinerja LSTM dengan *naive forecast*, ARIMA, GRU, Transformer, dan model hibrida melalui uji statistik terhadap perbedaan akurasi.

5. REFERENCES

- [1] Y. N. Muflikh, C. Smith, C. Brown, and A. A. Aziz, "Analysing price volatility in

- agricultural value chains using systems thinking: A case study of the Indonesian chilli value chain," *Agric. Syst.*, vol. 192, no. May, p. 103179, 2021, doi: 10.1016/j.agry.2021.103179.
- [2] B. Priambodo, Y. Jumaryadi, and U. Salamah, "Prediksi Perambatan Arus Lalu Lintas Berdasarkan Korelasi Tertinggi Antar Jalan," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 3, p. 667, 2022, doi: 10.30865/jurikom.v9i3.4203.
- [3] A. Wahab *et al.*, "Stock Prediction for Indonesia Stock Exchange with Long Short-Term Memory," *J. Ilm. FIFO*, vol. 16, no. 1, pp. 96–103, 2024.
- [4] G. Qin, X. Niu, J. Wang, and Q. Qian, "Traffic flow prediction based on JANet and attention mechanism," *2021 IEEE 6th Int. Conf. Intell. Comput. Signal Process. ICSP 2021*, no. Icsp, pp. 131–134, 2021, doi: 10.1109/ICSP51882.2021.9408777.
- [5] I. Nurhaida *et al.*, "Implementation of Deep Learning Predictor (LSTM) Algorithm for Human Mobility Prediction," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 18, pp. 132–144, 2020, doi: 10.3991/ijim.v14i18.16867.
- [6] L. Wiranda and M. Sadikin, "PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA," *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 8, no. 3, pp. 184–196, 2019, doi: 10.23887/JANAPATI.V8I3.19139.
- [7] M. David, I. Cholissodin, and N. Yudistira, "Prediksi Harga Cabai menggunakan Metode Long-Short Term Memory (Case Study : Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1214–1219, Mar. 2023.
- [8] T. Yue and Y. Liu, "Multi-Scale Price Forecasting Based on Data Augmentation," *Appl. Sci.*, vol. 14, no. 19, 2024, doi: 10.3390/app14198737.
- [9] A. Arfan and L. ETP, "Perbandingan Algoritma Long Short-Term Memory dengan SVR pada Prediksi Harga Saham di Indonesia," *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 13, no. 1, 2020, doi: 10.33322/petir.v13i1.858.
- [10] B. A. Abdelfattah, S. M. Darwish, and S. M. Elkaffas, "Enhancing the Prediction of Stock Market Movement Using Neutrosophic-Logic-Based Sentiment Analysis," *J. Theor. Appl. Electron. Commer. Res.*, vol. 19, no. 1, pp. 116–134, 2024, doi: 10.3390/jtaer19010007.
- [11] R. Mugi Setya Adi, "Prediksi Harga Komoditas Pangan Menggunakan Algoritma Long Short-Term Memory (LSTM)," *Technol. Sci.*, vol. 4, no. 2, 2022, doi: 10.47065/bits.v4i2.2229.
- [12] H. Suwoyo *et al.*, "Maximum likelihood estimation-assisted ASVSF through state covariance-based 2D SLAM algorithm," *Telkomnika (Telecommunication Comput. Electron. Control)*, vol. 19, no. 1, pp. 327–338, 2021, doi: 10.12928/TELKOMNIKA.V19I1.16223.
- [13] B. A. Memon, "Forecasting Fossil Energy Price Dynamics with Deep Learning: Implications for Global Energy Security and Financial Stability," *Algorithms*, vol. 18, no. 12, 2025, doi: 10.3390/a18120776.
- [14] N. Harshith and P. Kumari, "Memory based neural network for cumin price forecasting in Gujarat, India," *J. Agric. Food Res.*, vol. 15, no. January, p. 101020, 2024, doi: 10.1016/j.jafr.2024.101020.
- [15] C. Sun, M. Pei, B. Cao, S. Chang, and H. Si, "A Study on Agricultural Commodity Price Prediction Model Based on Secondary Decomposition and Long Short-Term Memory Network," *Agric.*, vol. 14, no. 1, 2024, doi: 10.3390/agriculture14010060.
- [16] S. H. Khan, M. Hayat, and F. Porikli, "Regularization of deep neural networks with spectral dropout," *Neural Networks*, vol. 110, pp. 82–90, 2019, doi: 10.1016/j.neunet.2018.09.009.