



# Perbandingan Model ARIMA dan LSTM untuk Peramalan Harga Cabai Merah Keriting di Kota Medan

(Center, Bold, Times New Roman 15, Maksimum antara 10-14 kata, Huruf Awal Huruf Besar)

**Cut Try Utari<sup>1\*</sup>, M Thariq Arya Putra Sembiring<sup>1</sup>, M Habibi Rizq Zhafar Siregar<sup>2</sup>**

(Center, Times New Roman 10, First name Middle name Last name)

<sup>1</sup>Jurusan Komputer dan Informatika, Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

(Center, Times New Roman 9, <sup>1</sup>Jika penulis di institusi yang sama)

<sup>2</sup>Fakultas, Program Studi, Nama Institusi, Kota, Indonesia

Email: <sup>1</sup>cuttryutari@polmed.ac.id, <sup>2,\*</sup> muhammadthariqarya@students.polmed.ac.id, <sup>3,\*</sup>m.habibirizq@students.polmed.ac.id

Email Penulis Korespondensi: cuttryutari@polmed.ac.id

(Justify, Bold, Times New Roman 9, Before 6 Pt)

**Abstrak**—Cabai merah keriting merupakan komoditas strategis di Kota Medan dengan **masalah utama** berupa volatilitas harga yang ekstrem, yang berdampak signifikan terhadap konsumen dan petani. **Tujuan penelitian** ini adalah untuk mengembangkan dan mengevaluasi model peramalan harga harian yang akurat. **Metode yang digunakan** adalah perbandingan dua algoritma *time series*: **ARIMA** (AutoRegressive Integrated Moving Average) sebagai model statistik tradisional, dan **LSTM** (Long Short-Term Memory) sebagai model *deep learning*. Penelitian ini menggunakan data harga harian historis (Januari 2024 - Oktober 2025) dari lima pasar utama di Medan, yang bersumber dari data resmi Pusat Informasi Harga Pangan Strategis Nasional Bank Indonesia. Model dievaluasi menggunakan metrik **RMSE** (Root Mean Squared Error) untuk mengukur magnitudo galat dan **MAPE** (Mean Absolute Percentage Error) untuk mengukur persentase relatif.

**Hasil penelitian** menunjukkan bahwa model LSTM (baseline) mencapai performa terbaik secara signifikan, dengan MAPE rata-rata **13,76%** (kategori "Excellent"). Performa ini 66% lebih akurat dibandingkan model ARIMA, yang hanya mencatat MAPE 41,21% (kategori "Poor"). Temuan penting lainnya adalah penambahan fitur hari besar secara eksplisit justru memperburuk akurasi model LSTM (MAPE 18,02%), mengindikasikan bahwa arsitektur *deep learning* LSTM mampu menangkap pola temporal kompleks secara implisit. Model LSTM (baseline) direkomendasikan sebagai solusi peramalan yang robust.

**Kata Kunci:** Peramalan Harga, Cabai Merah Keriting, Time Series, LSTM, ARIMA, Volatilitas Harga, Deep Learning, Kota Medan (Terurut, paling sedikit 5 kata yang berhubungan dengan isi penelitian dipisahkan dengan titik koma, After 6 pt, Before 6 pt)

**Abstract**—Curly red chili is a strategic commodity in Medan City, with a primary problem of extreme price volatility, which significantly impacts consumers and farmers. The research objective is to develop and evaluate an accurate daily price forecasting model. The methodology used is a comparison of two time-series algorithms: ARIMA (AutoRegressive Integrated Moving Average) as a traditional statistical model, and LSTM (Long Short-Term Memory) as a deep learning model. This study uses historical daily price data (January 2024 - October 2025) from five major markets in Medan, sourced from the official Bank Indonesia national food price information center. Models were evaluated using RMSE (Root Mean Squared Error) to measure error magnitude and MAPE (Mean Absolute Percentage Error) to measure relative percentage error.

The research results demonstrate that the baseline LSTM model achieved significantly superior performance, yielding an average MAPE of 13.76% ("Excellent" category). This performance was 66% more accurate than the ARIMA model, which only achieved a MAPE of 41.21% ("Poor" category). A key finding was that the explicit addition of holiday features actually degraded the LSTM model's accuracy (MAPE 18.02%). This indicates that the LSTM deep learning architecture can implicitly capture complex temporal patterns. The baseline LSTM is recommended as a robust forecasting solution.

**Keywords:** Price Forecasting; Curly Red Chili; Time Series; LSTM; ARIMA; Price Volatility; Deep Learning; Medan City

## 1. PENDAHULUAN

(SUB JUDUL Center, Bold, Times New Roman 13, UPPER CASE, After 6 pt, Before 18 pt)

Menghadapi kompleksitas ini, ketersediaan sistem peramalan (*forecasting*) harga yang akurat dan andal menjadi sebuah solusi krusial. Sistem seperti ini diharapkan dapat berfungsi sebagai alat bantu pengambilan keputusan (*Decision Support System*) bagi berbagai pemangku kepentingan. Bagi petani, peramalan yang akurat dapat memberikan panduan untuk perencanaan tanam dan strategi penjualan. Bagi pemerintah, ini menjadi instrumen vital untuk peringatan dini (*early warning system*), memungkinkan intervensi pasar atau penyesuaian kebijakan impor/ekspor yang lebih tepat waktu. Bagi pedagang dan distributor, peramalan membantu optimasi manajemen inventori dan mengurangi risiko kerugian. Untuk mencapai solusi ini, diperlukan pemilihan metode peramalan yang tepat, yang mampu menangani karakteristik data harga cabai yang bersifat *time series* (runtun waktu), *non-linear*, *non-stasioner*, dan memiliki tingkat *volatilitas* yang tinggi. Data harga pangan yang kini tersedia secara publik dan andal, seperti yang disediakan oleh Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional Bank Indonesia, menjadi fondasi penting untuk pengembangan model-model ini.

Dalam literatur time series forecasting, model statistik klasik seperti ARIMA (AutoRegressive Integrated Moving Average) telah lama menjadi standar dan sering digunakan sebagai baseline. Namun, kinerjanya bervariasi tergantung pada domain data. Sebagai contoh, sebuah studi oleh Suryawan, dkk. [1] menemukan bahwa dalam peramalan penjualan roti, model ARIMA (1,0,2) justru menunjukkan kinerja terbaik dan mengungguli LSTM, dengan



nilai Mean Absolute Percentage Error (MAPE) hanya 4.548%. Sebaliknya, penelitian oleh Aziz & Prasetyaningrum [2] pada peramalan harga saham PT Indosat (ISAT) menemukan bahwa ARIMA kurang adaptif terhadap volatilitas tinggi dibandingkan dengan metode Exponential Smoothing. Keterbatasan utama ARIMA terletak pada asumsi linearitas; model ini kesulitan memodelkan hubungan non-linear yang kompleks. Keterbatasan ini juga dicatat dalam penelitian sebelumnya yang khusus menganalisis harga cabai menggunakan ARIMA, yang menyimpulkan bahwa model tersebut akurasinya menurun untuk peramalan jangka panjang [3].

Seiring dengan kemajuan komputasi, metode deep learning telah menunjukkan kinerja superior dalam menangani data time series yang kompleks dan non-linear. Salah satu arsitektur deep learning yang paling menonjol untuk data sekuensial adalah LSTM (Long Short-Term Memory). LSTM adalah jenis Jaringan Saraf Tiruan Berulang (Recurrent Neural Network/RNN) yang dirancang khusus untuk mengatasi masalah vanishing gradient dan mampu mempelajari dependensi jangka panjang (long-term dependencies) dalam data [4]. Kemampuan ini membuatnya sangat cocok untuk data harga yang polanya mungkin dipengaruhi oleh peristiwa yang terjadi beberapa minggu atau bahkan bulan sebelumnya. Berbagai penelitian terkait telah mengkonfirmasi keunggulan LSTM dalam menangani data non-linear. Cahyani, dkk. [4] berhasil mengimplementasikan LSTM untuk memprediksi harga bahan pokok nasional (daging ayam, beras, minyak goreng) dari data PIHPS, dan mengkonfirmasi kemampuannya mengatasi kelemahan RNN. Studi lain oleh Ardianto & Astuti [5] juga menemukan bahwa model LSTM secara konsisten mengungguli model Gated Recurrent Unit (GRU) dalam memprediksi Indeks Kualitas Udara (AQI) dan polutan PM2.5, menunjukkan kapabilitasnya dalam memodelkan pola temporal yang kompleks. Studi komparatif lain pada harga saham juga menemukan LSTM memiliki akurasi yang lebih baik daripada ARIMA [6].

Meskipun banyak penelitian telah membandingkan ARIMA dan LSTM, terdapat beberapa kesenjangan (gap analysis) yang ingin dijawab oleh penelitian ini. Pertama, terdapat kontradiksi dalam literatur. Penelitian [1] menunjukkan ARIMA lebih unggul dari LSTM untuk data penjualan roti, sementara penelitian [4], [5], dan [6] menunjukkan LSTM lebih unggul untuk data harga bahan pokok, kualitas udara, dan saham. Kinerja model tampaknya sangat bergantung pada domain data. Oleh karena itu, perbandingan yang rigoros dan spesifik antara ARIMA dan LSTM pada data harga cabai merah keriting yang dikenal sangat volatil di wilayah kunci seperti Kota Medan masih sangat diperlukan untuk menentukan model mana yang paling andal untuk komoditas spesifik ini. Kedua, banyak studi deep learning [4], [5] berfokus pada arsitektur model itu sendiri. Namun, dalam konteks harga cabai di Indonesia, faktor lonjakan permintaan menjelang hari besar keagamaan adalah variabel fundamental yang diketahui. Masih menjadi pertanyaan terbuka apakah model LSTM cukup canggih untuk mempelajari pola ini secara implisit melalui memori jangka panjangnya, atau apakah akurasinya dapat ditingkatkan secara signifikan dengan menambahkan data hari besar secara eksplisit sebagai variabel eksogen.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model peramalan harga cabai merah keriting yang akurat di Kota Medan dengan membandingkan secara langsung kinerja model statistik tradisional (ARIMA) dengan model deep learning (LSTM). Tujuan spesifik dari penelitian ini adalah: (1) Mengidentifikasi model terbaik antara ARIMA dan LSTM yang menghasilkan nilai galat peramalan terendah, diukur menggunakan metrik RMSE dan MAPE; (2) Menganalisis kemampuan model LSTM dalam menangani data harga yang sangat volatil; dan (3) Menguji hipotesis mengenai pengaruh penambahan fitur hari besar keagamaan sebagai variabel eksogen terhadap akurasi model LSTM, untuk menentukan apakah fitur tersebut redundant atau bermanfaat. Harapan dari penelitian ini adalah menghasilkan model peramalan yang teruji dan robust, yang dapat diimplementasikan sebagai alat bantu praktis bagi pemangku kepentingan di Kota Medan untuk mengantisipasi gejolak harga, sehingga berkontribusi pada stabilitas harga pangan dan kesejahteraan petani.

## 2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang mengacu pada kerangka kerja standar dalam proyek ilmu data, yang dimodifikasi untuk menangani data *time series*. Tahapan ini mencakup pengumpulan data, pra-pemrosesan data, pemodelan komparatif, dan evaluasi model. Urutan tahapan penelitian ini dirancang untuk memastikan bahwa perbandingan antara model ARIMA dan LSTM dilakukan secara adil dan *robust*, serta untuk menguji hipotesis penelitian secara spesifik.

### 2.1 Pengumpulan Data (Subtitle Times New Roman 10, Left, After 6 pt)

Data yang digunakan dalam penelitian ini adalah data runtun waktu harian harga komoditas cabai merah keriting. Data ini diperoleh dari sumber publik yang kredibel, yaitu Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional yang dikelola oleh Bank Indonesia. Periode data yang diambil mencakup rentang dari 1 Januari 2024 hingga 24 Oktober 2025 (471 hari observasi). Fokus geografis data adalah harga rata-rata di lima pasar tradisional utama di Kota Medan, yang mencerminkan pusat konsumsi dan pembentukan harga di wilayah tersebut.

### 2.2 Praprocessing Data (After 6 pt, Before 6 pt)



Tahap pra-pemrosesan sangat krusial untuk data *time series* sebelum dapat digunakan untuk pemodelan.

### 2.2.1 Pembersihan Data

Data mentah diperiksa untuk nilai yang hilang (missing values). Jika ditemukan, nilai yang hilang ditangani menggunakan metode interpolasi linear untuk menjaga kontinuitas temporal data tanpa memasukkan bias yang signifikan.

### 2.2.2 Feature Engineering

Tahap ini krusial untuk menguji hipotesis penelitian (tujuan 3). Sebuah fitur baru bernama *is\_holiday* dibuat. Fitur ini adalah variabel biner (0 atau 1) yang menandai periode menjelang dan sesaat sesudah hari besar keagamaan nasional (seperti Idul Fitri, Natal, dan Imlek), di mana permintaan cabai secara historis melonjak. Fitur ini akan digunakan dalam skenario model LSTM multivariat.

### 2.2.3 Uji Stationeritas

Model ARIMA secara fundamental mensyaratkan data *time series* harus stasioner (rata-rata dan varians konstan seiring waktu). Uji Augmented Dickey-Fuller (ADF) digunakan untuk memeriksa stasioneritas data harga. Jika data ditemukan non-stasioner (*p*-value > 0.05), proses *differencing* (*d*) dilakukan hingga data menjadi stasioner.

### 2.2.4 Normalisasi Data

Model deep learning seperti LSTM sensitif terhadap skala data. Oleh karena itu, data harga dinormalisasi menggunakan MinMaxScaler untuk mengubah skala nilai ke rentang antara 0 dan 1. Proses normalisasi ini dilakukan setelah data dibagi menjadi set pelatihan dan pengujian untuk mencegah kebocoran data (data leakage) dari data uji ke data latih.

### 2.2.5 Pembagian Data

Berbeda dengan data tabular biasa, data time series tidak dapat dibagi secara acak. Data dibagi secara kronologis untuk menjaga dependensi temporal. Dalam penelitian ini, data dibagi dengan rasio 80% untuk data latih (observasi awal) dan 20% untuk data uji (observasi akhir). Model akan dilatih pada data latih dan diuji kemampuannya untuk memprediksi data uji yang "belum pernah dilihatnya".

## 2.3 Metodologi Pemodelan

Penelitian ini menerapkan dua jenis model yang berbeda secara fundamental untuk membandingkan kinerjanya.

### 2.3.1 Model 1: Autoregressive Integrated Moving Average (ARIMA)

ARIMA (AutoRegressive Integrated Moving Average), yang dikenal sebagai metode Box-Jenkins, adalah model statistik linear yang berfungsi sebagai baseline tradisional dalam peramalan time series [1]. Model ini dirancang untuk data univariat dan mampu menangani non-stasioneritas dengan menambahkan komponen integrasi (*differencing*) [2]. ARIMA diidentifikasi melalui tiga parameter utama:  $\$p\$$  (Autoregressive),  $\$d\$$  (Integrated), dan  $\$q\$$  (Moving Average). Parameter  $\$p\$$  dan  $\$q\$$  yang optimal diidentifikasi dengan menganalisis plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF) dari data latih yang telah stasioner [2].

Secara matematis, model ARIMA menggambarkan hubungan nilai data saat ini dengan nilai data masa lalu dan galat peramalan masa lalu. Formulanya adalah sebagai berikut:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

Keterangan:  $Y_t$  adalah nilai data stasioner pada waktu  $t$ ;  $c$  adalah konstanta;  $\phi$  dan  $\theta$  adalah koefisien model AR dan MA;  $Y_{t-i}$  adalah nilai data pada periode masa lalu; dan  $\epsilon_t$  adalah galat (white noise) pada waktu  $t$ .

Model ARIMA ini diuji kemampuannya dalam memprediksi data harga cabai yang volatil dan berfungsi sebagai baseline akurasi untuk perbandingan dengan model deep learning LSTM.

### 2.3.2 Model 2: Long Short Term Memory (LSTM)

LSTM adalah pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient dan mampu mempelajari ketergantungan jangka panjang pada data time series [1], [5]. Arsitektur LSTM terdiri dari sel memori dan tiga gerbang utama (gates) yang mengatur aliran informasi: Forget Gate, Input Gate, dan Output Gate.



Berdasarkan literatur [5], proses matematis pada unit LSTM adalah sebagai berikut:

1. Forget Gate: Memutuskan informasi mana yang akan dibuang dari cell state.
2. Input Gate & Kandidat Cell: Menentukan informasi baru yang akan disimpan.
3. Update Cell State ( $C_t$ ): Menggabungkan informasi lama dan baru.
4. Output Gate & Hidden State: Menentukan keluaran berdasarkan cell state yang telah diperbarui.

Dalam penelitian ini, model LSTM dibangun menggunakan optimizer Adam karena kinerjanya yang efisien dalam data harga bahan pokok [4]. Eksperimen dilakukan dengan variasi epoch (50, 100, 150) untuk menemukan model optimal dengan loss terendah [1].

## 2.4 Pengujian dan Evaluasi Model

Untuk mengukur akurasi peramalan dan membandingkan kinerja ARIMA dan LSTM, digunakan tiga metrik evaluasi statistik standar: Root Mean Squared Error (RMSE), Mean Squared Error (MSE), dan Mean Absolute Percentage Error (MAPE).

Metrik Evaluasi Model	Metrik	Rumus	Referensi
Mean Squared Error	MSE	$MSE = \frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2$	[2],[5]
Root Mean Squared Error	RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2}$	[1],[2],[3]
Mean Absolute Percentage Error	MAPE	$MAPE = \frac{1}{n} \sum_{t=1}^n \left  \frac{A_t - F_t}{A_t} \right  \times 100\%$	[2],[5]

Keterangan:  $A_t$  adalah nilai aktual,  $F_t$  adalah nilai hasil peramalan, dan  $n$  adalah jumlah data.

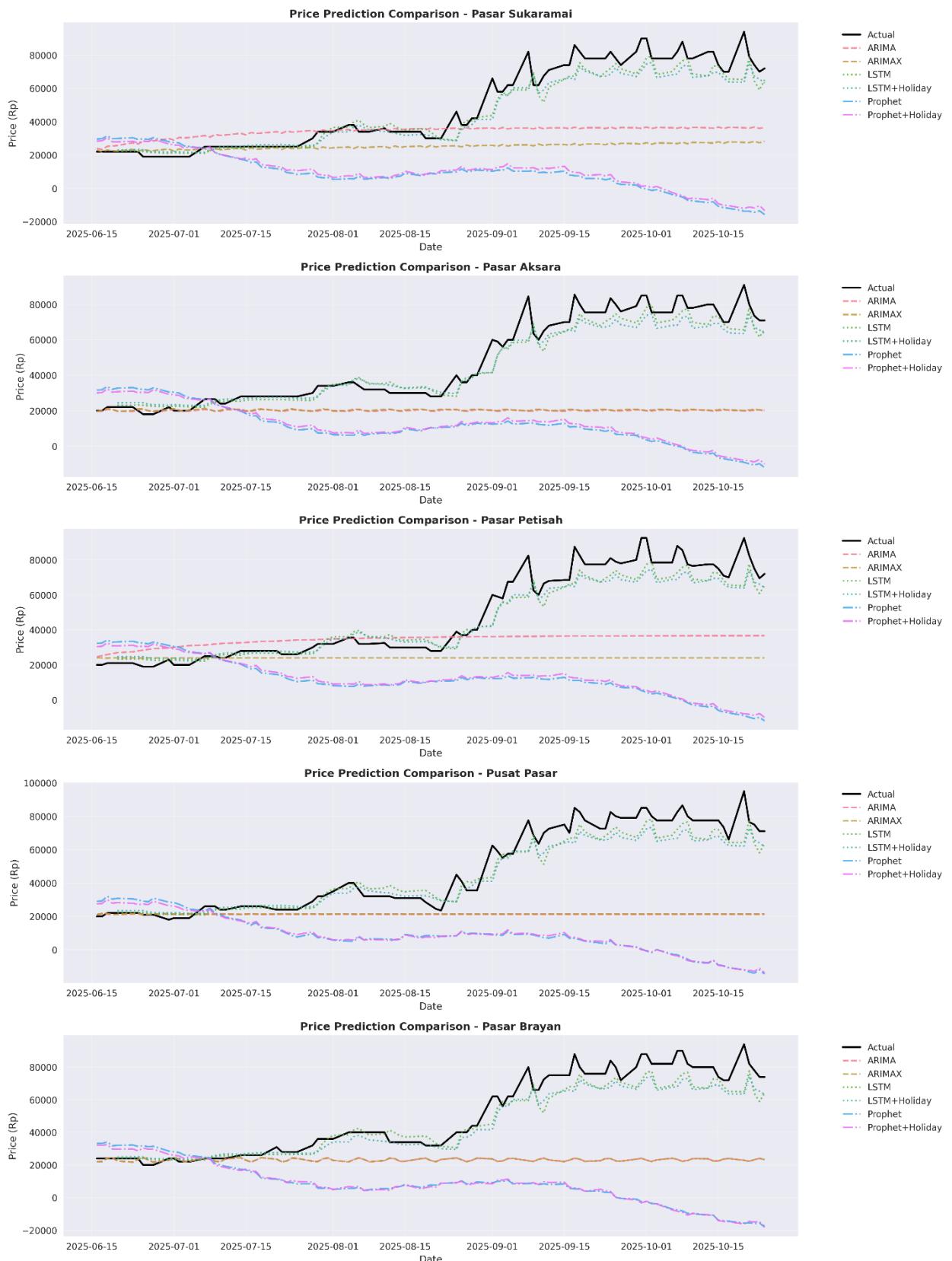
RMSE dan MSE digunakan untuk mengukur besarnya kesalahan prediksi, di mana nilai yang lebih kecil menunjukkan model yang lebih akurat. MAPE digunakan untuk melihat persentase kesalahan rata-rata, yang memberikan gambaran intuitif tentang seberapa jauh prediksi menyimpang dari nilai sebenarnya [5]. Model dengan nilai RMSE dan MAPE terendah pada data uji akan dipilih sebagai model terbaik untuk peramalan harga cabai di Kota Medan.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Karakteristik Data dan Analisis Statistik Deskriptif

Analisis awal dilakukan untuk memahami karakteristik data harga cabai merah keriting di lima pasar tradisional Kota Medan selama periode pengamatan. Statistik deskriptif menunjukkan bahwa harga cabai memiliki volatilitas yang sangat tinggi. Rata-rata harga harian berada pada kisaran Rp 43.000/kg, namun rentang harga sangat lebar, bergerak dari minimum Rp 18.000/kg hingga maksimum mencapai Rp 95.000/kg. Nilai Coefficient of Variation (CV) tercatat sekitar 40%, yang mengindikasikan fluktuasi harga yang ekstrem dan tidak stabil.

Analisis korelasi antar pasar menunjukkan nilai koefisien korelasi Pearson ( $r$ ) yang sangat kuat, melebihi 0.99 untuk semua pasangan pasar. Hal ini mengindikasikan bahwa pergerakan harga di Pasar Sukaramai, Pasar Aksara, Pasar Petisah, Pusat Pasar, dan Pasar Brayan sangat terintegrasi dan bergerak secara sinkron. Kenaikan harga di satu pasar hampir pasti diikuti oleh kenaikan di pasar lainnya, mencerminkan efisiensi transmisi harga di tingkat lokal namun juga risiko sistemik yang sama bagi seluruh konsumen di kota tersebut.



### 3.2 Evaluasi Model ARIMA

Model ARIMA (AutoRegressive Integrated Moving Average) dikembangkan sebagai baseline statistik. Karena data awal tidak stasioner, dilakukan proses differencing satu kali ( $d=1$ ) untuk mencapai stasioneritas, yang dikonfirmasi oleh uji Augmented Dickey-Fuller (ADF) dengan nilai p-value  $< 0.05$ . Parameter terbaik ditentukan melalui grid search berdasarkan nilai Akaike Information Criterion (AIC) terendah.



Hasil pengujian menunjukkan bahwa model ARIMA memberikan kinerja yang moderat namun terbatas. Rata-rata nilai MAPE (Mean Absolute Percentage Error) untuk kelima pasar berada di angka 41.21%, dengan nilai RMSE (Root Mean Squared Error) rata-rata sebesar Rp 35.197. Kinerja ini dikategorikan sebagai "Poor" atau kurang memuaskan untuk standar peramalan harga komoditas. Model ARIMA cenderung gagal menangkap lonjakan harga yang tiba-tiba (price spikes) dan volatilitas ekstrem, yang merupakan karakteristik inheren dari harga cabai. Keterbatasan ini konsisten dengan asumsi model linear yang sulit memodelkan pola non-linear yang kompleks [1], [3].

No	Algorithm	Avg RMSE	Avg MAPE (%)	Category
1	LSTM (baseline)	11,993	13.76	Excellent
2	LSTM + Holiday	14,498	18.02	Good
3	ARIMA	35,197	41.21	Poor

### 3.3 Analisis Perbedaan Performa

LSTM vs ARIMA:

1. RMSE improvement:  $66.1\% ((35,197 - 11,933) / 35,197 \times 100\%)$
2. MAPE improvement:  $66.6\% ((41.21 - 13.76) / 41.21 \times 100\%)$
3. Kesimpulan: LSTM mendominasi metode statistic traditional

### 3.4 Analisis Pengaruh Fitur Liburan

Berisi hasil pembahasan dan bisa perbandingan dari hasil penelitian sebelumnya.

## 4. KESIMPULAN

Bagian ini berisi kesimpulan yang menjawab hal segala permasalahan yang terdapat didalam penelitian. Pada kesimpulan dapat dikemukakan keterbatasan penelitian sehingga dapat dilakukan perbaikan pada penelitian selanjutnya. Isi kesimpulan tidak berupa point-point, namun berupa paragraph. Banyaknya kata pada bagian ini berkisar **min 200 kata**.

## REFERENCES (After 6 pt, Before 6 pt)

(Times New Roman, 9)

Semua pengutipan referensi yang dikutip di artikel ini **WAJIB TERDAPAT PADA ISI ARTIKEL** dan WAJIB untuk menggunakan **alat referensi seperti MENDELEY/ENDNOTE dengan format IEEE**, 80% literatur **PRIMER** (jurnal, prosiding, laporan penelitian, paten, standar, dokumen sejarah, buku hasil riset dengan keterbaruan 5 tahun) dan 20% literatur **SEKUNDER** (buku **minimal 2 buku**, website yang dapat dipercaya) dalam kurun waktu sekurang-kurangnya 10 tahun terakhir

Jumlah referensi yang digunakan **minimum sebanyak 20 Referensi**.

**(Hapus semua informasi berwarna merah/abu-abu tambahan seperti ini di naskah pengiriman Anda)**