

Implementasi Metode *Clustering* dan *Denoising* dalam Peramalan Deret Waktu Data Harga Bahan Pangan di Indonesia dengan Model SARIMA

Nailfaaz
Departemen Ilmu Komputer dan
Elektronika
Universitas Gadjah Mada
Yogyakarta, Indonesia
nailfaaz@mail.ugm.ac.id

Muhammad Dafa Wisnu Galih
Departemen Ilmu Komputer dan
Elektronika
Universitas Gadjah Mada
Yogyakarta, Indonesia
dafa.w.dev@gmail.com

Krisna Bayu Dharma Putra
Departemen Ilmu Komputer dan
Elektronika
Universitas Gadjah Mada
Yogyakarta, Indonesia
dharna.competition@gmail.com

Abstract—Fluktuasi harga bahan pangan di Indonesia menjadi tantangan dalam menjaga stabilitas ekonomi dan ketahanan pangan. Penelitian ini bertujuan untuk meramalkan harga 13 komoditas pangan dengan menggunakan model Seasonal Autoregressive Integrated Moving Average (SARIMA), yang mampu menangkap pola musiman pada data time series. Data diperoleh dari kompetisi Kaggle Datavidia 9.0 yang mencakup periode 1 Januari 2022 hingga 30 September 2024. Pendekatan yang digunakan meliputi clustering untuk mengklasifikasikan komoditas berdasarkan volatilitas harga, serta denoising menggunakan Gaussian Wavelet Denoising guna mengurangi noise dan meningkatkan akurasi peramalan. Proses imputasi missing values dilakukan dengan metode backward fill, Dynamic Time Warping (DTW), dan linear interpolation untuk memastikan kelengkapan data. Hasil klusterisasi menunjukkan bahwa Cabai Rawit Merah memiliki volatilitas tertinggi, sementara Daging Sapi Murni tergolong paling stabil. Evaluasi model menunjukkan bahwa SARIMA memberikan hasil prediksi yang lebih akurat dibandingkan ARIMA, terutama pada komoditas dengan tren musiman. Penerapan fine-tuning berhasil menurunkan nilai Mean Absolute Percentage Error (MAPE) rata-rata sebesar 19,1%, dengan penurunan paling signifikan pada Bawang Merah (45,47% → 7,01%) dan Cabai Merah Keriting (18,62% → 4,91%).

Keywords—Forecasting, harga bahan pangan, SARIMA, denoising, clustering, volatilitas, MAPE.

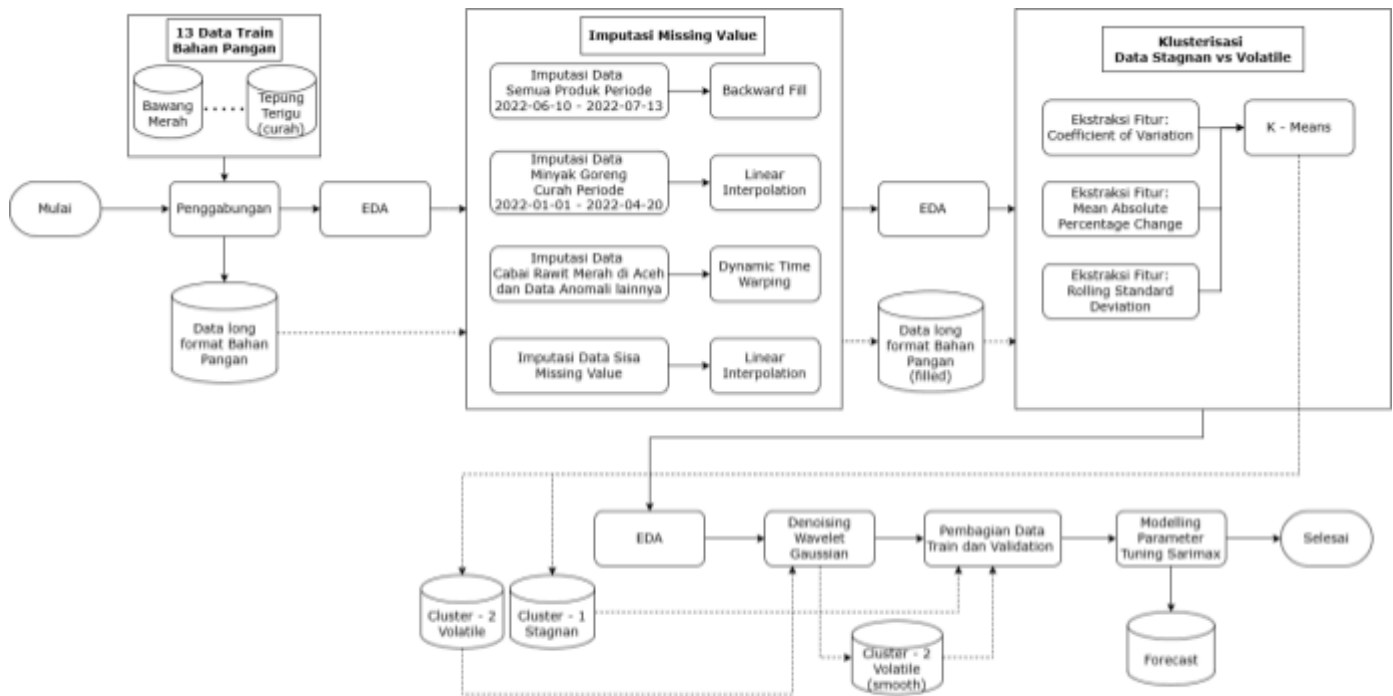
I. PENDAHULUAN

Ketahanan pangan merupakan sebuah isu yang menjadi perhatian global, terutama dalam menjaga keseimbangan antara ketersediaan pangan dan keterjangkauan harga. Ketahanan pangan sendiri didefinisikan sebagai kondisi di mana individu dalam suatu negara dapat mengakses pangan dalam jumlah dan kualitas yang memadai untuk memenuhi kebutuhan gizi mereka [1], [2]. Berdasarkan laporan Global Food Security Index (GFSI), Indonesia saat ini menempati peringkat ke-63 dari 113 negara dalam indeks ketahanan pangan global [3]. Salah satu aspek yang masih menjadi

perhatian adalah ketersediaan pangan, yang tercermin dari skor pilar Availability yang hanya mencapai 50.9 poin. Ketersediaan pangan yang kurang optimal dapat memicu kenaikan harga bahan pangan, terutama ketika pasokan tidak mampu memenuhi permintaan yang terus meningkat [4]. Lonjakan harga ini dapat berdampak pada daya beli masyarakat, khususnya bagi kelompok ekonomi menengah ke bawah, sehingga memperburuk ketahanan pangan nasional. Oleh karena itu, langkah strategis diperlukan untuk mengendalikan harga dan memastikan stabilitas pasar pangan di Indonesia.

Salah satu langkah preventif yang dapat dilakukan untuk menjaga stabilitas harga pangan di pasaran adalah melalui pengondisian harga dengan berbagai kebijakan, seperti regulasi harga, subsidi, dan operasi pasar [5]. Pemerintah dapat menerapkan intervensi harga untuk menghindari lonjakan harga yang tidak terkendali, misalnya dengan menetapkan harga acuan atau mengatur distribusi bahan pangan agar lebih merata [6]. Selain itu, pengelolaan stok pangan nasional juga menjadi strategi penting dalam menjaga kestabilan harga di pasaran. Namun, efektivitas dari langkah-langkah ini bergantung pada momentum yang tepat dalam implementasinya. Pemilihan waktu yang kurang tepat dapat menyebabkan kebijakan menjadi kurang efektif atau bahkan berdampak sebaliknya, seperti kelangkaan pasokan yang justru memperburuk kondisi pasar. Oleh karena itu, diperlukan metode yang dapat membantu dalam menentukan waktu terbaik untuk melakukan intervensi harga.

Salah satu cara untuk menentukan waktu yang tepat dalam pengondisian harga pangan adalah dengan melakukan analisis tren serta peramalan atau forecasting harga pangan di pasaran. Peramalan harga pangan berperan dalam memberikan wawasan mengenai kemungkinan fluktuasi harga di masa mendatang sehingga pengambilan keputusan dapat lebih berbasis data [7]. Metode forecasting ini erat kaitannya dengan model statistik serta perkembangan teknologi yang memungkinkan analisis data dalam skala besar. Dengan kemajuan dalam bidang machine learning (ML) dan deep learning (DL), model-model prediktif semakin berkembang untuk meningkatkan akurasi peramalan harga pangan. Berbagai metode statistik seperti regresi, time series analysis, hingga model berbasis ML dan DL seperti RNN, LSTM, dan Transformer telah diterapkan



Gambar 1. Diagram Alir Penelitian secara Keseluruhan

untuk memahami pola harga komoditas pangan secara lebih mendalam [8]-[10].

Pada penelitian kali ini, fokus diberikan pada peramalan harga pangan untuk 13 jenis komoditas di pasaran. Peramalan dilakukan dengan menganalisis tipe tren harga dari setiap komoditas, apakah bersifat fluktuatif atau cenderung stagnan. Selain itu, dilakukan proses denoising untuk mengurangi noise pada data guna menghasilkan nilai peramalan yang lebih stabil dan akurat [11]. Model yang digunakan dalam penelitian ini adalah Seasonal Autoregressive Integrated Moving Average (SARIMA), yang dikenal memiliki performa baik dalam menangkap pola musiman pada data time series [12]. Dengan pendekatan ini, penelitian bertujuan untuk memberikan hasil prediksi harga pangan yang nantinya diharapkan dapat digunakan sebagai salah satu pertimbangan dalam kebijakan pengendalian harga serta upaya stabilisasi ketahanan pangan nasional.

II. METODOLOGI

Sebelum membahas detail metode yang akan digunakan dalam penelitian, Gambar 1. memvisualisasikan bagaimana penelitian ini berjalan secara keseluruhan melalui diagram alir.

Penelitian dimulai dengan menggabungkan 13 jenis data bahan pangan menjadi satu set data terintegrasi dan dieksplorasi melalui proses EDA (Exploratory Data Analysis) untuk memahami pola, distribusi, dan karakteristik awal. Setelah itu, dilakukan imputasi nilai hilang dengan beberapa metode, seperti *backward fill*, *dynamic time warping*, dan *linear interpolation* guna memastikan data lengkap. Data yang sudah diimputasi

kemudian diklasifikasikan menjadi *stagnant* atau *volatile* dengan menggunakan ekstraksi fitur dan metode K-Means.

Khusus untuk data dengan karakteristik volatilitas tinggi, dilakukan proses *denoising* menggunakan wavelet Gaussian agar sinyal utama data lebih terlihat. Selanjutnya, data dibagi menjadi set latih dan set validasi, lalu dimodelkan menggunakan proses parameter tuning untuk mendapatkan performa terbaik. Model yang sudah optimal digunakan untuk melakukan peramalan (*forecast*) pergerakan harga bahan pangan pada periode mendatang, sehingga dihasilkan output akhir berupa prediksi harga untuk submisi di Kaggle dan analisis lebih lanjut.

A. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari folder train Harga Bahan Pangan pada kompetisi Kaggle. Folder ini berisi 13 file, masing-masing mewakili satu jenis komoditas, yaitu Bawang Merah, Bawang Putih Bonggol, Beras Medium, Beras Premium, Cabai Merah Keriting, Cabai Rawit Merah, Daging Ayam Ras, Daging Sapi Murni, Gula Konsumsi, Minyak Goreng Curah, Minyak Goreng Kemasan Sederhana, Telur Ayam Ras, dan Tepung Terigu (Curah).

Sebagai sampel, dataset Bawang Merah.csv digunakan untuk inspeksi awal. Dataset ini memiliki 35 kolom, terdiri dari 1 kolom timestamp (Date) dan 34 kolom harga yang mewakili masing-masing provinsi di Indonesia. Data ini mencakup 1.004 baris dengan rentang waktu dari 1 Januari 2022 hingga 30 September 2024. Cuplikan data dalam format CSV dapat dilihat pada Tabel I.

Setelah dilakukan pengecekan terhadap seluruh dataset dalam folder train, ditemukan bahwa setiap file memiliki karakteristik yang seragam, yaitu dimensi yang sama,

rentang tanggal yang sama, dan nama kolom yang sama. Keseragaman ini memungkinkan penggabungan semua dataset menjadi satu file, sehingga analisis dapat dilakukan dengan lebih efisien.

TABEL I. CUPLIKAN DATA HARGA BAHAN PANGAN (BAWANG MERAH.CSV)

Date	Aceh	Bali	Banten	Sumatera Utara
2022-01-01	28970	20870	26890	28710
2022-01-02	29900	20710	25600	28460
2022-01-03	28970	20510	26390	28050
....
2024-09-30	28960	24230	27940	27150

Seluruh dataset dalam folder train telah digabungkan menjadi satu file dengan nama `train_merged.csv`, yang terdiri dari 4 kolom utama: Date (tanggal pencatatan), Region (nama provinsi), Price (harga bahan pangan), dan Product (nama produk pangan). Dataset hasil penggabungan ini memiliki 443.768 baris, yang merupakan hasil dari kombinasi 34 provinsi, 13 produk, dan 1.004 hari. Format ini disajikan dalam bentuk long format, yang lebih fleksibel untuk eksplorasi dan pemodelan. Cuplikan data long format ini dapat dilihat pada Tabel II.

TABEL II. CUPLIKAN DATA MODIFIKASI DENGAN LONG FORMAT

Date	Region	Price	Product
2022-01-01	Aceh	28970	Bawang Merah
2022-01-02	Aceh	29900	Bawang Merah
2022-01-03	Aceh	28970	Bawang Merah
....
2024-09-30	Sumatera Utara	10640	Tepung Terigu (Curah)

Sebagai catatan, dataset test tidak digabungkan karena dalam proses peramalan harga, model hanya memerlukan data train sebagai dasar untuk melakukan prediksi.

B. Metode Imputasi Missing Value

Dalam penelitian ini, imputasi diperlukan untuk mengisi *missing value* dalam dataset. Beberapa metode yang digunakan meliputi interpolasi linear, backward fill, dan Dynamic Time Warping (DTW).

1. Interpolasi Linear.

Metode ini digunakan untuk mengisi nilai yang hilang dengan memperkirakan nilai di antara dua titik waktu berdasarkan tren linear. Dengan metode ini, nilai-nilai yang hilang diisi berdasarkan tren linear antara dua titik yang tersedia. Rumus interpolasi linear adalah:

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{(x_2 - x_1)}$$

Dengan:

- y adalah nilai yang diinterpolasi.

- x adalah titik waktu yang hilang.
- x_1, x_2 adalah dua titik waktu sebelum dan sesudah *missing value*.
- y_1, y_2 adalah nilai harga pada x_1 dan x_2 .

2. Backward fill.

Metode ini mengisi nilai yang hilang dengan menggunakan nilai dari titik waktu setelahnya. Metode ini efektif untuk mengisi data yang hilang pada awal rentang waktu, dengan mengasumsikan bahwa tren harga cenderung berkelanjutan dalam waktu dekat. Secara matematis, dapat dirumuskan sebagai:

$$y_t = \begin{cases} y_{t+1}, & \text{jika } y_t \text{ hilang} \\ y_t, & \text{jika } y_t \text{ tersedia} \end{cases}$$

Dengan:

- y adalah nilai yang diinterpolasi.
- y_t adalah nilai pada waktu t .
- Jika y_t hilang, diisi dengan nilai dari waktu y_{t+1} .
- Jika y_t tersedia, maka nilainya tetap.

3. Dynamic Time Warping (DTW).

Algoritma ini digunakan untuk mengukur kemiripan antara dua deret waktu, bahkan jika terdapat perbedaan dalam kecepatan atau panjang waktu. Metode ini bekerja dengan menghitung jarak optimal antara dua urutan waktu dengan cara membandingkan titik-titik dalam pola yang mungkin tidak sejajar secara langsung. Secara matematis, DTW dirumuskan sebagai:

$$DTW(X, Y) = \min \sum_{i=1}^n d(x_i, y_j)$$

Dengan:

- $X = (x_1, x_2, \dots, x_n)$ adalah deret waktu yang dibandingkan dengan $Y = (y_1, y_2, \dots, y_n)$.
- $d(x_i, y_j)$ adalah jarak antara dua titik data.
- Algoritma mencari jalur optimal dalam matriks jarak untuk meminimalkan perbedaan total..

C. Variabel Clustering

Dalam penelitian ini, sebanyak tiga indikator pembantu digunakan untuk mengklasterisasi pola harga ke dalam dua kategori utama: *volatile* dan *stagnan*. Beberapa indikator statistik yang digunakan yaitu Coefficient of Variation (CV), Mean Absolute Percentage Change (MAPC), dan Rolling Standard Deviation.

1. Coefficient of Variation (CV).

Variabel ini mengukur seberapa besar variasi harga dibandingkan dengan rata-ratanya. Secara matematis dapat dirumuskan sebagai:

$$CV = \frac{\sigma}{\mu}$$

Dengan:

- σ adalah standar deviasi harga dalam rentang waktu tertentu.
- μ adalah rata-rata harga dalam rentang waktu tersebut.
- Semakin tinggi nilai CV, semakin besar fluktuasi harga relatif terhadap rata-ratanya, yang menunjukkan tingkat volatilitas yang lebih tinggi.

2. Mean Absolute Percentage Change (MAPC).

Metode ini mengukur seberapa besar perubahan harga secara relatif dari satu periode ke periode berikutnya. Rumusnya adalah:

$$MAPC = \frac{1}{n} \sum_{t=1}^n \left| \frac{P_t - P_{t-1}}{P_{t-1}} \right| \times 100\%$$

Dengan:

- P_t adalah harga pada waktu t .
- P_{t-1} adalah harga pada periode sebelumnya.
- Semakin tinggi nilai MAPC, semakin besar perubahan harga yang terjadi dari waktu ke waktu, yang menunjukkan tingkat volatilitas yang lebih tinggi.

3. Rolling Standard Deviation.

Metode ini mengukur fluktuasi harga dalam jendela waktu tertentu, dengan cara menghitung standar deviasi harga dalam periode berjalan. Secara matematis dapat dirumuskan sebagai berikut:

$$Rolling\ Std = \sqrt{\frac{1}{n} \sum_{i=t}^{t+n} (P_i - \bar{P})^2}$$

Dengan:

- P_t adalah harga dalam *window* waktu tertentu.
- \bar{P} adalah rata-rata harga dalam *window* waktu tersebut yang digunakan sebagai titik acuan.
- Semakin besar rolling standard deviation, semakin besar fluktuasi harga dalam periode tertentu, yang mengindikasikan volatilitas pasar yang lebih tinggi.

D. K-Means Clustering

Setelah indikator atau variabel statistik yang menggambarkan volatilitas harga dihitung, kami menerapkan K-Means Clustering untuk mengelompokkan data ke dalam dua kluster utama: Kluster Stagnan dan Kluster Volatile.

K-Means adalah salah satu algoritma klusterisasi yang bekerja dengan membagi data ke dalam sejumlah kelompok berdasarkan kemiripan nilai fitur. Algoritma ini secara iteratif mencari titik pusat kluster (centroid) yang optimal, sehingga setiap data dalam satu kluster memiliki karakteristik yang lebih mirip satu sama lain dibandingkan

dengan data di kluster lain. Secara matematis, formulanya sebagai berikut:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

Dengan:

- k adalah jumlah kluster (dalam hal ini, 2 kluster: stagnan dan volatile).
- C_i adalah kluster ke- i .
- P_t adalah centroid dari kluster ke- i .
- x adalah data yang termasuk ke dalam kluster tersebut.

E. Gaussian Wavelet Denoising

Gaussian Wavelet Denoising adalah satu metode yang digunakan untuk denoising, yaitu teknik pemrosesan data time series yang memungkinkan penyaringan noise serta ekstraksi pola utama dalam data. Metode ini didasarkan pada fungsi Gaussian yang telah diturunkan beberapa kali untuk menghasilkan berbagai level transformasi. Dengan menggunakan pendekatan ini, fluktuasi kecil yang bersifat acak dapat dihilangkan, sementara pola utama tetap dipertahankan.

Secara matematis, fungsi dasar dari Wavelet Gaussian dirumuskan sebagai berikut:

$$\psi(t) = -\frac{2}{\sqrt{3}\sigma\pi^{\frac{1}{4}}} \left(1 - \frac{t^2}{\sigma^2}\right) e^{-\frac{t^2}{2\sigma^2}}$$

Dengan:

- $\Psi(t)$ adalah fungsi wavelet Gaussian.
- σ adalah parameter skala yang menentukan lebar gelombang.
- t adalah variabel waktu atau posisi dalam data.

F. Seasonal Autoregressive Integrated Moving Average

SARIMA adalah metode pemodelan time series yang menggabungkan komponen ARIMA (Autoregressive Integrated Moving Average) dengan faktor musiman (seasonality). Model ini sangat berguna untuk data yang menunjukkan pola berulang dalam interval waktu tertentu, seperti data bulanan atau tahunan. Dengan menambahkan komponen musiman, SARIMA dapat menangkap pola musiman yang tidak bisa dijelaskan oleh ARIMA biasa.

Secara umum, model SARIMA dinotasikan sebagai:

$$SARIMA(p, d, q) \times (P, D, Q, s)$$

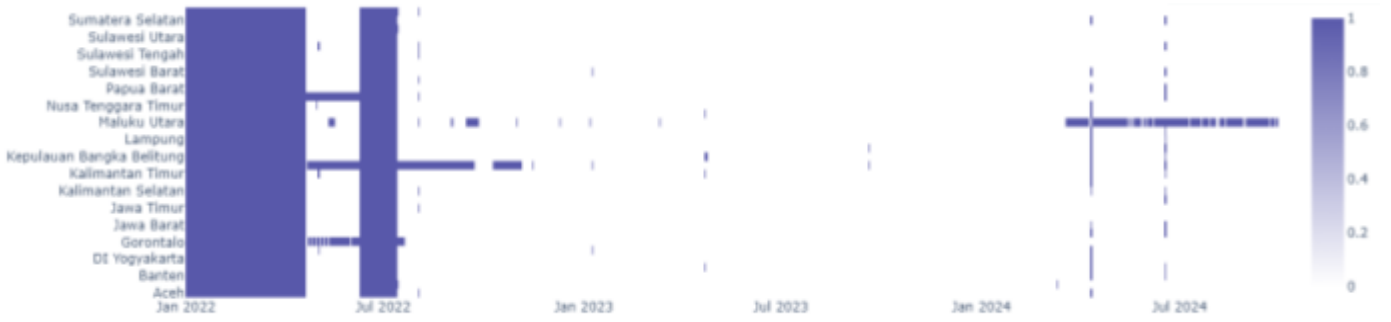
dengan parameter musiman (P, D, Q, s) yang memperhitungkan pola musiman dalam data.

Secara matematis, SARIMA dapat dirumuskan dalam bentuk berikut:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^DY_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t$$

Dengan:

- Y_t adalah nilai time series pada waktu t .



Gambar 2. Missing Value pada Produk Minyak Goreng Curah

- B adalah operator backshift, dengan $B^k Y_t = Y_{t-k}$
- $\phi_p(B)$ adalah polinomial autoregressive nonmusiman orde p .
- $\Phi_P(B^s)$ adalah polinomial autoregressive musiman orde P dengan periode s .
- $(1 - B)^d$ adalah operator differencing nonmusiman orde d .
- $(1 - B^s)^d$ adalah operator differencing musiman orde D dengan periode s .
- $\theta_q(B)$ adalah polinomial moving average nonmusiman orde q .
- $\Theta_Q(B^s)$ adalah polinomial autoregressive musiman orde Q dengan periode s .
- ϵ_t adalah noise white noise dengan distribusi normal.

SARIMA pada dasarnya merupakan ekstensi dari ARIMA yang mencakup faktor musiman. Jika parameter musiman $(P, D, Q, s) = (0, 0, 0, 0)$, maka SARIMA akan kembali menjadi ARIMA biasa.

Dengan kata lain, ARIMA adalah kasus khusus dari SARIMA ketika tidak ada komponen musiman dalam data. Model SARIMA sering digunakan dalam analisis time series dengan pola musiman.

G. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah salah satu metrik evaluasi yang digunakan untuk mengukur seberapa besar kesalahan dalam prediksi dibandingkan dengan nilai aktualnya dalam bentuk persentase. MAPE sering digunakan dalam pemodelan time series dan regresi karena memberikan gambaran intuitif tentang tingkat kesalahan relatif terhadap skala data.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%$$

Dengan:

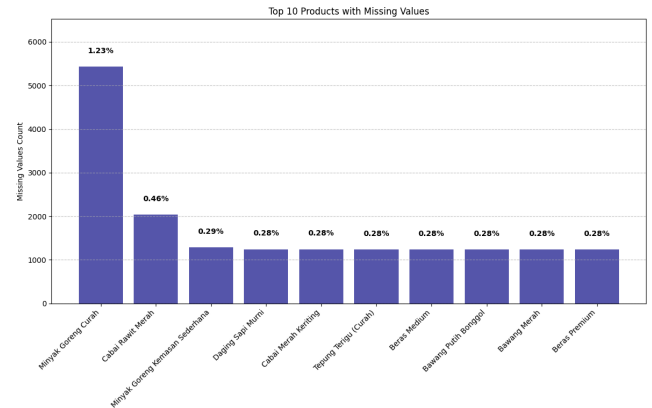
- Y_t adalah nilai aktual pada waktu t .

- \hat{Y}_t adalah nilai prediksi pada waktu t .
- n adalah jumlah total observasi.

III. EXPLORATORY DATA ANALYSIS

A. Analisis Missing Value

Dalam analisis data harga bahan pangan, keberadaan missing value, Gambar. 2, merupakan tantangan utama yang perlu ditangani sebelum dilakukan proses pemodelan dan peramalan. Data yang hilang dapat mengurangi akurasi model dan menyebabkan bias dalam analisis tren harga. Oleh karena itu, langkah awal yang dilakukan adalah mengidentifikasi pola missing value di setiap kombinasi provinsi dan produk guna memahami distribusi serta penyebab potensial dari hilangnya data.



Gambar 3. Persentase Missing Value pada Top 9 Produk

Berdasarkan inspeksi data, ditemukan bahwa setiap produk di setiap provinsi mengalami missing value dengan pola yang cukup konsisten. Normalnya, jumlah missing value berkisar antara 34 hingga 39 setelah dilakukan agregasi berdasarkan Provinsi + Produk, Lampiran. 1. Namun, terdapat beberapa anomali signifikan yang perlu diperhatikan:

1. Missing Value Tinggi produk Minyak Goreng Curah
 - Pada hampir setiap provinsi, jumlah missing value untuk Minyak Goreng Curah jauh lebih besar dibandingkan produk lainnya, berkisar antara 100 hingga 300 data yang hilang.
 - Ditemukan pola missing value yang terkonsentrasi di awal periode pencatatan, Gambar. 2, yakni dari

1 Januari 2022 hingga 20 April 2022.

2. Missing Value Ekstrem di Cabai Rawit Merah Provinsi Aceh

- Produk Cabai Rawit Merah di Aceh memiliki jumlah missing value tertinggi di antara semua kombinasi provinsi + produk, Lampiran 2.
- Ditemukan 782 data yang hilang, yang setara dengan 77% dari total baris untuk produk ini di provinsi tersebut.

3. Pola Konsisten: Rentang Waktu Missing Value

Semua produk menunjukkan pola missing value yang konsisten dalam rentang 10 Juni 2022 hingga 13 Juli 2022, Lampiran 3.

4. Data Anomali Lainnya

Selain dua kasus utama di atas, terdapat beberapa produk dan provinsi lain yang memiliki jumlah missing value lebih tinggi dari batas normal (>39 baris), terutama di luar rentang anomali yang sudah disebutkan. Beberapa kasus mencolok meliputi Cabai Rawit Merah di Aceh dan Sumatera Barat, Minyak Goreng Curah di Gorontalo, Maluku Utara, Kalimantan Utara, dan Papua, serta Minyak Goreng Kemasan Sederhana di Gorontalo.

B. Kevolatilitasan

Dapat dilihat pada Lampiran 3, secara visual tampak jelas bahwa terdapat banyak garis data yang menunjukkan volatilitas tinggi. Untuk menganalisis fenomena ini secara lebih mendalam secara statistik, kami mengekstraksi nilai agregat dari CV, MAPC, dan Rolling Standard Deviation guna mengukur tingkat fluktuasi harga secara kuantitatif.

TABEL III. HASIL EKSTRAKSI FITUR

Product	CV	MAPC	Rolling_Std
Bawang Merah	0.193	0.006	423
Bawang Putih Bonggol	0.166	0.003	143
Beras Medium	0.09	0.001	21
Beras Premium	0.09	0.001	21
Cabai Merah Keriting	0.20	0.01	1035
Cabai Rawit Merah	0.20	0.01	1306
Daging Ayam ras	0.03	0.003	179
Daging Sapi Murni	0.02	0.002	309
Gula Konsumsi	0.09	0.001	24
Minyak Goreng Curah	0.10	0.003	58
Minyak Goreng Kemasan Sederhana	0.11	0.003	74
Telur Ayam Ras	0.06	0.003	120
Tepung Terigu (Curah)	0.05	0.002	22

Berdasarkan hasil ekstraksi fitur volatilitas yang ditampilkan dalam Tabel III, dapat disimpulkan bahwa terdapat perbedaan yang signifikan dalam tingkat fluktuasi

harga antar produk pangan.

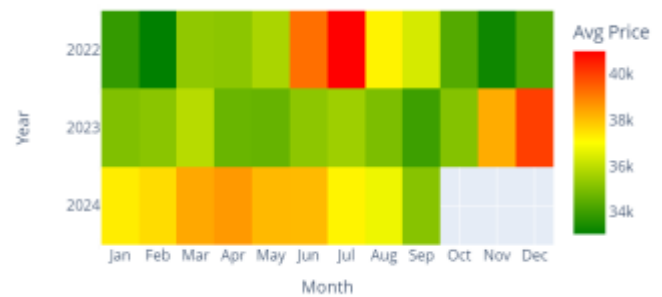
Produk-produk seperti Cabai Rawit Merah (CV: 0.20, MAPC: 0.01, Rolling Std: 1306) dan Cabai Merah Keriting (CV: 0.20, MAPC: 0.01, Rolling Std: 1035) memiliki nilai tertinggi dalam semua metrik, menandakan bahwa harga kedua produk ini mengalami fluktuasi yang sangat tinggi dari waktu ke waktu.

Sebaliknya, produk-produk seperti Daging Sapi Murni (CV: 0.02, MAPC: 0.002, Rolling Std: 309) dan Daging Ayam Ras (CV: 0.03, MAPC: 0.003, Rolling Std: 179) menunjukkan volatilitas yang jauh lebih rendah, menandakan bahwa harga produk-produk ini lebih stabil dalam jangka waktu tertentu.

Selain itu, bahan pokok seperti Beras Medium, Beras Premium, Gula Konsumsi, dan Tepung Terigu (Curah) memiliki nilai CV yang rendah (0.05 - 0.09) serta Rolling Std yang kecil, menegaskan bahwa fluktuasi harga produk-produk ini relatif terkendali.

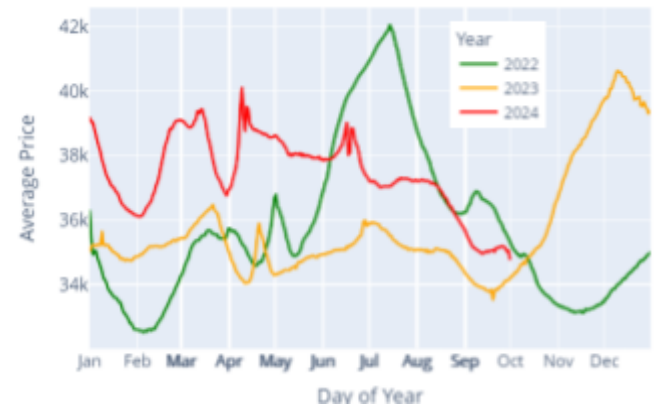
C. Seasonality dan Tren

Pada bagian ini, kami menganalisis pola seasonality dan tren harga bahan pangan untuk mengidentifikasi pola tahunan atau musiman yang dapat menjadi pertimbangan dalam tahap modeling. Analisis dilakukan dari skala luas (tahun dan bulan) hingga lebih mendetail (hari dan bulan).



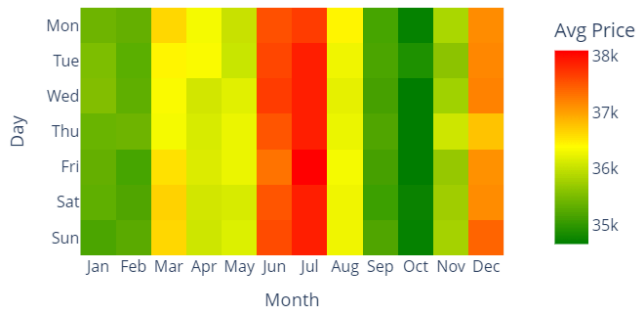
Gambar 4. Heatmap Rerata Harga Pangan (Tahun vs Bulan)

Gambar 4. menunjukkan bahwa harga tertinggi terjadi pada Juli 2022 dan Desember 2023, ditandai dengan warna merah, sementara harga terendah terlihat di Februari dan November 2022. Pada tahun 2024, meskipun tidak ada perubahan rata-rata yang signifikan, harga cenderung lebih tinggi secara keseluruhan, terlihat dari warna kuning-oranye yang mendominasi.



Gambar 5. Tren Rerata Harga Bahan Pangan per Tahun

Selanjutnya Gambar 5. menggambarkan tren tahunan, di mana harga cenderung rendah pada Februari, September, Oktober, dan November, sementara kenaikan signifikan terjadi di bulan puasa dan menjelang Idul Fitri (Maret–Mei), sebelum kembali turun dan mengalami kenaikan dalam waktu yang singkat. Selain itu, terdapat tren kenaikan harga di akhir tahun (Desember).



Gambar 6. Heatmap Rerata Harga Pangan (Hari vs Bulan)

Lalu, Gambar 6. menunjukkan bahwa harga pangan cenderung lebih mahal pada pertengahan dan akhir tahun, serta lebih murah di awal tahun. Secara mingguan, terdapat sedikit kenaikan harga saat akhir pekan, terutama di bulan Maret dan Desember, di mana harga pada Sabtu dan Minggu sedikit lebih tinggi dibanding hari lainnya.

Berdasarkan beberapa visualisasi dan analisis mengenai seasonality dan tren ini, dapat ditarik beberapa poin *insight*:

- Tahun 2024 menunjukkan harga lebih tinggi secara keseluruhan, kemungkinan dipengaruhi oleh inflasi.
- Terdapat pola seasonality terkait momen Idul Fitri, Idul Adha, dan akhir tahun.
- Selain pola tahunan, terdapat seasonality bulanan, di mana bulan tertentu memiliki harga lebih tinggi.
- Model yang mampu mengenali pola *seasonality* dapat memberikan prediksi yang lebih akurat.

IV. HASIL DAN PEMBAHASAN

A. Imputasi Missing Value

Dalam menangani missing value pada dataset ini, digunakan beberapa metode imputasi yang disesuaikan dengan pola hilangnya data serta karakteristik produk di setiap provinsi. Pemilihan metode dilakukan berdasarkan jumlah missing value, posisi waktu hilangnya data, serta kemungkinan adanya keterkaitan dengan tren harga di wilayah lain.

1. Interpolasi Linear

Untuk sebagian besar produk dengan missing value yang tidak terlalu besar dan tersebar dalam kurun waktu tertentu (10 Juni 2022 hingga 13 Juli 2022), digunakan metode interpolasi linear. Metode ini mengisi nilai yang hilang berdasarkan tren harga sebelum dan sesudahnya, sehingga nilai yang diimputasi tetap mengikuti pola perubahan harga yang ada.

2. Backward Fill untuk Minyak Goreng Curah

Produk Minyak Goreng Curah memiliki pola missing value yang unik, di mana data awal (1 Januari 2022 hingga

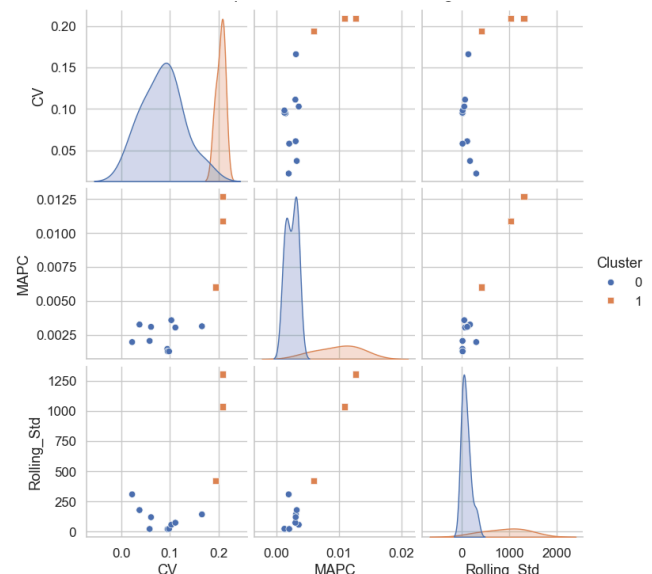
20 April 2022) tidak tersedia di hampir semua provinsi. Karena tidak ada data sebelumnya untuk digunakan dalam interpolasi, metode Backward Fill diterapkan.

3. DTW Produk Khusus dan Anomali Lain

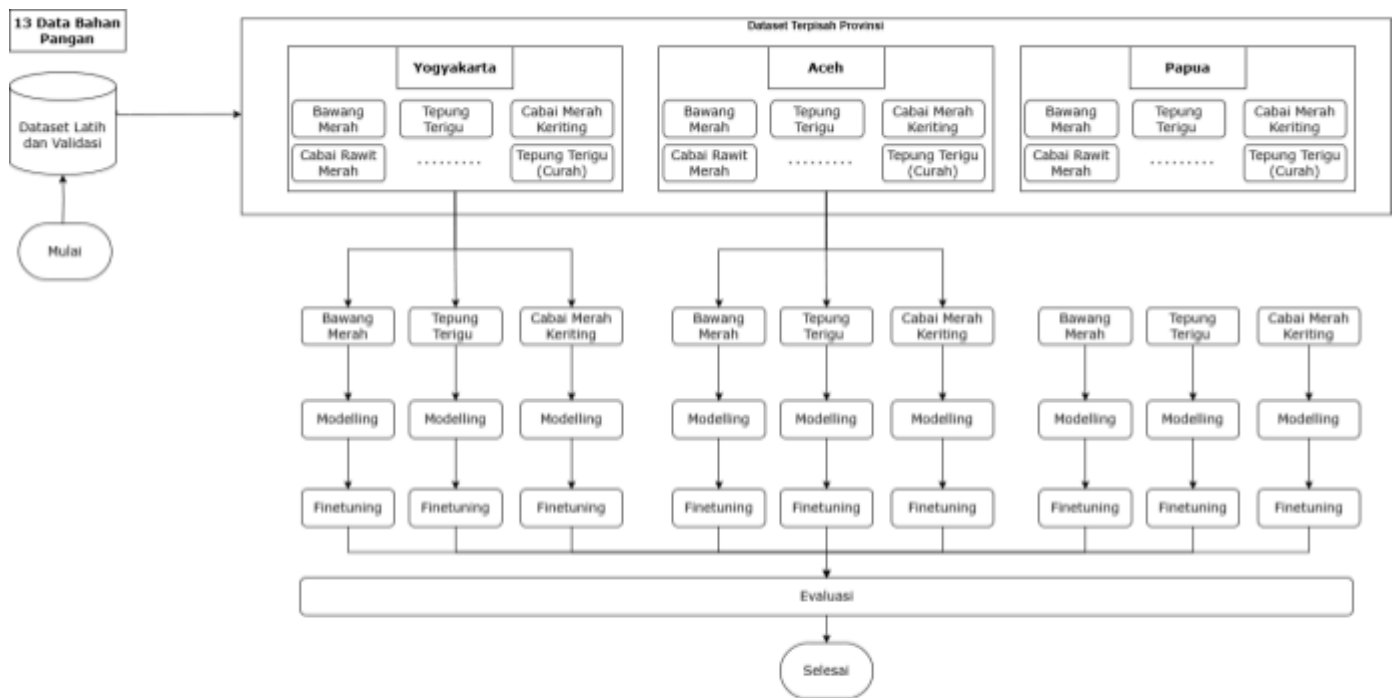
Untuk kasus khusus seperti di produk Cabai Rawit Merah di Aceh dan beberapa anomali lainnya, interpolasi sederhana atau backward fill tidak cukup efektif karena risiko mengisi data dengan pola yang tidak representatif. Oleh karena itu, digunakan pendekatan DTW, yang memungkinkan pencocokan tren harga dengan provinsi lain yang memiliki pola harga paling mirip. Dengan cara ini, nilai yang hilang diisi berdasarkan referensi dari wilayah lain yang memiliki pergerakan harga yang sebanding.

B. Products Clustering

Setelah melakukan ekstraksi fitur CV, MAPC, dan Rolling Standard Deviation, dilakukan proses clustering menggunakan algoritma K-Means dengan jumlah kluster $K = 2$. Tujuan dari pendekatan ini adalah untuk secara otomatis dan matematis mengelompokkan setiap produk ke dalam salah satu dari dua kluster, yaitu stagnan atau volatile, berdasarkan pola pergerakan harga historisnya. Hasil dari klusterisasi dapat dilihat pada Gambar. 7 dan Tabel IV. Secara visual, Gambar. 8, hasil klusterisasi juga jelas terlihat perbedaannya



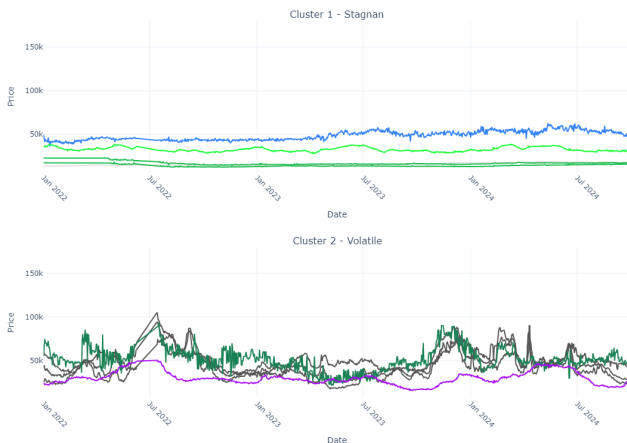
Gambar 7. Visualisasi K-Means dengan Fitur CV, MAPC, dan Rolling Standard Deviation



Gambar 8. Diagram Alir Modeling

TABEL IV. HASIL KLASTERISASI K-MEANS(K=2)

Kluster 1 - stagnan	Kluster 2 - volatile
Bawang Putih Bonggol	Cabai Rawit Merah
Minyak Goreng Kemasan Sederhana	Cabai Merah Keriting
Daging Ayam Ras	Bawang Merah
Minyak Goreng Curah	
Telur Ayam Ras	
Beras Premium	
Beras Medium	
Gula Konsumsi	
Tepung Terigu (Curah)	
Daging Sapi Murni	



Gambar 8. Visualisasi Perbedaan Klasterisasi Data Stagnan dan Volatile

C. Price Denoising

Dilakukan metode lebih lanjut untuk data yang tergolong volatile, dengan menerapkan smoothing menggunakan algoritma wavelet Gaussian. Metode ini diterapkan dengan parameter db dan level 2, yang bertujuan untuk mereduksi fluktuasi harga yang bersifat noise tanpa menghilangkan pola utama dalam data. Salah satu hasil dari metode ini dapat dilihat pada Gambar 9.



Gambar 9. Visualisasi Metode Wavelet Gaussian pada Data Bawang Merah Kalimantan Timur

Hasil smoothing ini diharapkan dapat menyediakan data yang lebih stabil dan representatif bagi proses peramalan harga. Dengan menghilangkan gangguan dari lonjakan harga yang bersifat anomali atau berskala kecil, model prediktif dapat lebih fokus pada tren musiman dan pola harga jangka panjang.

D. Forecast Modelling

1. Splitting, Modelling, dan Finetuning

Pelatihan diawali dengan membagi dataset menjadi dua bagian, yaitu data latih dan data uji seperti pada Gambar 8. Setelah itu, dataset akan dibagi berdasarkan provinsinya. Jadi setiap provinsi akan memiliki 13 jenis komoditas. Lalu,

tiap-tiap komoditas akan dipisahkan satu sama lain sehingga dataset berupa sebuah data komoditas per provinsi per komoditas. Setiap data komoditas ini akan independen dengan komoditas lain, baik yang memiliki provinsi sama maupun provinsi yang berbeda.

Masing-masing data komoditas akan memasuki fase *modelling* secara independen dengan dua model yang akan dipakai adalah SARIMA dan ARIMA. Model SARIMA akan lebih cocok untuk data yang memiliki tren dan *seasonality*, sedangkan ARIMA baik pada data yang cenderung tidak memiliki tren maupun *seasonality*.

Model terbaik dari masing-masing pelatihan akan memasuki masa *finetuning*, di mana akan dicari hyperparameter terbaik dari model tersebut untuk dataset komoditas tertentu pada provinsi tertentu.

Untuk menentukan model mana yang akan di-*finetuning*, dilakukan percobaan pada 3 buah komoditas dengan tingkat volatilitas yang tinggi pada lima buah provinsi dengan rentang waktu dari bulan Juli 2024 hingga bulan September 2024. Tiga buah komoditas tersebut adalah Bawang Merah, Cabai Rawit Merah, dan Cabai Merah Keriting. Tiga buah provinsi yang diambil sebagai sampel adalah Kalimantan Utara, Kalimantan Timur, dan Kalimantan Tengah.

Pada percobaan pertama, dilakukan perbandingan model *baseline* ARIMA dengan SARIMA dengan hasil sesuai dengan Tabel V.

TABEL V. KOMPARASI MODEL ARIMA DENGAN SARIMA

Komoditas	Rata-rata MAPE ARIMA	Rata-rata MAPE SARIMA
Bawang Merah	42,57%	45,47%
Cabai Rawit Merah	22,24%	14,55%
Cabai Merah Keriting	32,20%	18,62%

Tabel V. menunjukkan hasil perbedaan pemodelan dengan ARIMA dan SARIMA pada ketiga buah komoditas. Hasil pelatihan menunjukkan bahwa pada komoditas Bawang Merah, ARIMA memiliki keunggulan dibandingkan dengan SARIMA sebesar 2,9%. Di lain sisi, model SARIMA menunjukkan performa yang mengesankan pada pemodelan Cabai Rawit Merah dan Cabai Merah Keriting dengan MAPE sebesar 14,55% dan 18,62% secara berurutan, mengungguli model ARIMA sebesar 7,69% dan 13,58%.

Selain itu, hasil analisis menunjukkan bahwa harga ketiga komoditas pada bulan Juli hingga September 2024, menunjukkan kestabilan dan tidak terlalu fluktuatif. Ketiga komoditas ini menunjukkan kenaikan harga ketika ada momen-momen penting, seperti tahun baru. Oleh karena itu, model SARIMA yang memiliki kemampuan menangkap tren dan *seasonality* lah yang dipilih sebagai model akhir untuk di-*tuning* dengan harapan model ini mampu menangkap tren dan *seasonality* harga pada komoditas yang akan diprediksi.

2. Forecasting

Setelah didapatkan model akhir untuk setiap komoditas, akan dilakukan *finetuning*. Proses *finetuning* dilakukan untuk setiap komoditas per region, sehingga setiap komoditas di setiap region akan memiliki konfigurasi hyperparameter yang berbeda satu sama lain. Hyperparameter yang akan di-*tuning* adalah nilai p, d, q yang mewakili *order*, serta P, D, Q yang mewakili *seasonality*.

TABEL VI. VARIABEL TUNING PARAMETER SARIMA

Hyperparameter	Nilai
p	[0,1,2]
d	[0,1]
q	[0,1,2]
P	[0,1]
D	[0,1]
Q	[0,1]

Tabel VI menunjukkan konfigurasi dari hyperparameter yang mungkin. Konfigurasi terbaik akan dicari dengan menggunakan *gridsearch* untuk setiap kombinasi. Kombinasi terbaik yang menghasilkan MAPE terendah akan dipilih sebagai model akhir untuk komoditas pada provinsi tersebut.

TABEL VII. SAMPEL EVALUASI MODEL SARIMA

Komoditas	MAPE Sebelum Finetuning	MAPE Setelah Finetuning
Bawang Merah	45,47%	7,01%
Cabai Rawit Merah	14,55%	9,35%
Cabai Merah Keriting	18,62%	4,91%

Perbandingan hasil setelah dan sebelum *finetuning* dapat dilihat pada Tabel VII. Hasil perbandingan menunjukkan bahwa metode *finetune* berhasil meningkatkan kinerja model secara signifikan. Pada komoditas cabai rawit merah, terjadi penurunan nilai MAPE sebesar 5,2%. Penurunan juga terjadi pada komoditas cabai merah keriting dengan tingkat penurunan sebesar 13,71%. Di lain sisi, penurunan nilai MAPE terbesar dialami oleh bawang merah, dengan nilai penurunan sebesar 38,46%.

Secara rata-rata, metode *finetuning* berhasil menurunkan MAPE model SARIMA pada ketiga jenis komoditas pada 3 provinsi Kalimantan Tengah, Kalimantan Utara, dan Kalimantan Timur sebesar 19,1%. Hal ini menunjukkan bahwa metode MAPE efektif untuk meningkatkan kemampuan model Sarima dalam memahami pola serta trend harga komoditas pasar.

V. KESIMPULAN

Penelitian ini mengimplementasikan metode clustering dan denoising untuk meningkatkan akurasi peramalan harga bahan pangan di Indonesia menggunakan model SARIMA. Data dari 13 komoditas pangan dikumpulkan dan diolah, dengan missing values diatasi menggunakan interpolasi linear, backward fill, dan Dynamic Time Warping (DTW). Hasil klasterisasi menggunakan K-Means menunjukkan bahwa Cabai Rawit Merah memiliki volatilitas tertinggi dengan CV sebesar 0,20, MAPC 0,01, dan Rolling Std 1306, diikuti oleh Cabai Merah Keriting dengan CV 0,20, MAPC 0,01, dan Rolling Std 1035, sementara Daging Sapi Murni memiliki volatilitas paling rendah dengan CV 0,02, MAPC 0,002, dan Rolling Std 309.

Denoising menggunakan Gaussian Wavelet Denoising berhasil mengurangi noise dan mempertajam pola harga sebelum dilakukan peramalan dengan SARIMA. Hasil evaluasi menunjukkan bahwa SARIMA lebih unggul dibandingkan ARIMA, terutama untuk komoditas yang memiliki tren musiman. Perbandingan model menunjukkan bahwa MAPE pada Cabai Rawit Merah turun dari 14,55% menjadi 9,35% setelah fine-tuning, sedangkan Cabai Merah Keriting mengalami penurunan dari 18,62% menjadi 4,91%, dan Bawang Merah turun signifikan dari 45,47% menjadi 7,01%. Secara keseluruhan, metode fine-tuning berhasil menurunkan MAPE rata-rata sebesar 19,1%, menunjukkan peningkatan performa model yang signifikan.

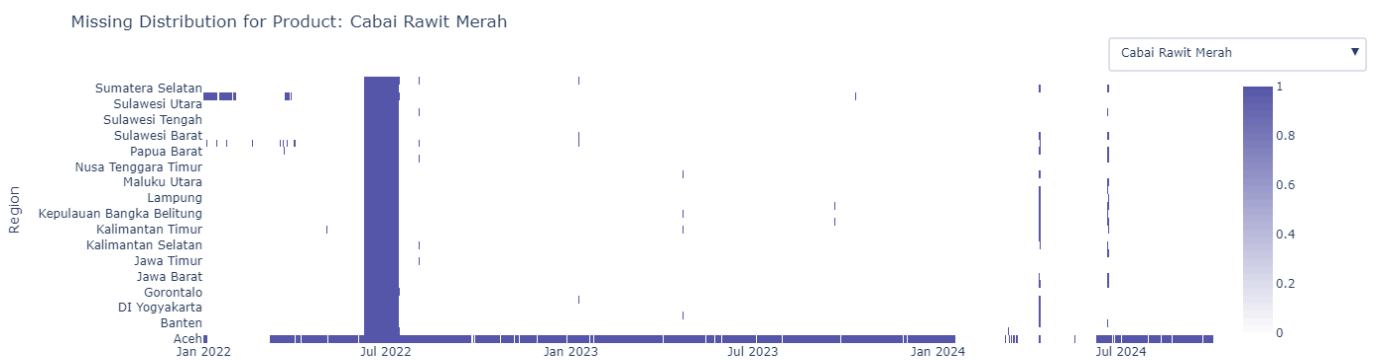
Hasil penelitian ini dapat digunakan sebagai dasar dalam pengambilan keputusan terkait kebijakan harga bahan pangan di Indonesia, khususnya dalam mengantisipasi lonjakan harga dan menjaga stabilitas ekonomi.

DAFTAR PUSTAKA

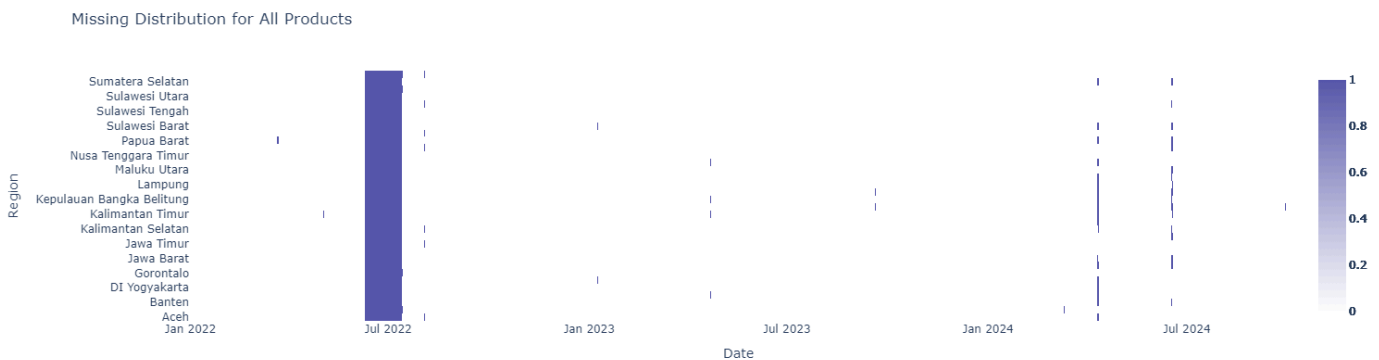
- [1] Indonesia. (2012). *Undang-Undang Tentang Pangan Nomor 18 Tahun 2012*. Diakses: 15 Maret 2025. [Online]. Tersedia: <https://peraturan.bpk.go.id/Home/Details/39100>.
- [2] Food and Agriculture Organization, "Report of the World Food Summit," World Food Summit - Final Report - Part 1, <https://www.fao.org/3/w3548e/w3548e00.htm#Medrano>, tanggal akses: 15 Maret 2025.
- [3] (2022) Global Food Security Index website. [Online], <https://impact.economist.com/sustainability/project/food-security-index/>, tanggal akses: 15 Maret 2025.
- [4] Universitas Minneapolis, *Principles of Economics*. Minneapolis, MN, Amerika Serikat: Open Textbook Library, 2016.
- [5] "Volume Impor (Berat Bersih) Menurut Golongan Penggunaan Barang (Ribu Ton), 2020-2022," Badan Pusat Statistik, <https://www.bps.go.id/indicator/8/1837/1/volume-impor-berat-bersih-menurut-golongan-penggunaan-barang.html>, tanggal akses: 15 Maret 2025.
- [6] Indonesia, Presiden. (2022). *Peraturan Presiden Nomor 125 tentang Penyelenggaraan Cadangan Pangan Pemerintah*. [Online]. Tersedia: <https://peraturan.bpk.go.id/Home/Details/230367/perpres-no-125-tahun-2022>.
- [7] L. Zeng, L. Ling, D. Zhang, and W. Jiang, "Optimal Forecast Combination based on PSO-CS approach for daily agricultural future prices forecasting," *Applied Soft Computing*, vol. 132, 2023. doi:10.1016/j.asoc.2022.109833.
- [8] Y.-S. Lee and W.-Y. Liu, "Forecasting value of agricultural imports using a novel two-stage hybrid model," *Computers and Electronics in Agriculture*, vol. 104, pp. 71–83, 2014. doi:10.1016/j.compag.2014.03.011.
- [9] S. M. Nejad *et al.*, "Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 254–266, 2023. doi:10.1109/jstars.2022.3223423.
- [10] S. Prakash, A. S. Jalal, and P. Pathak, "Forecasting covid-19 pandemic using Prophet, LSTM, Hybrid Gru-LSTM, CNN-LSTM, Bi-LSTM and stacked-LSTM for India," *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, 2023. doi:10.1109/iscon57294.2023.10112065.
- [11] Y. Ning, H. Kazemi, and P. Tahmasebi, "A comparative machine learning study for time series oil production forecasting: Arima, LSTM, and prophet," *Computers & Geosciences*, vol. 164, p. 105126, 2022. doi:10.1016/j.cageo.2022.105126.
- [12] B. A. Sunjaya, S. D. Permai, and A. A. Gunawan, "Forecasting of covid-19 positive cases in Indonesia using long short-term memory (LSTM)," *Procedia Computer Science*, vol. 216, pp. 177–185, 2023. doi:10.1016/j.procs.2022.12.125.

Product	Bawang Merah	Bawang Putih	Bonggol	Beras Medium	Beras Premium	Cabai Merah Keriting	Cabai Rawit Merah	Daging Ayam Ras	Daging Sapi Murni	Gula Konsumsi	Minyak Goreng Curah	Minyak Goreng Kemasan Sederhana	Telur Ayam Ras	Tepung Terigu (Curah)
Region														
Aceh	37		37	37	37	37	762	37	37	37	147		37	37
Bali	36	36	36	36	36	39	36	36	36	36	146		36	36
Banten	37	37	37	37	37	37	37	37	37	37	147		37	37
Bengkulu	37	37	37	37	37	37	37	37	38	37	148		37	37
DI Yogyakarta	36	36	36	36	36	36	36	36	36	36	146		36	36
DKI Jakarta	37	37	37	37	37	37	37	37	37	37	148		37	37
Gorontalo	35	35	35	35	35	35	35	35	35	35	193		82	35
Jambi	38	38	38	38	38	38	38	38	38	38	148		38	38
Jawa Barat	37	37	37	37	37	37	37	37	37	37	147		37	37
Jawa Tengah	34	34	34	34	34	34	34	34	34	34	144		34	34
Jawa Timur	35	35	35	35	35	35	35	35	35	35	145		35	35
Kalimantan Barat	36	36	36	36	36	36	36	36	36	36	146		36	36
Kalimantan Selatan	37	37	37	37	37	37	37	37	38	37	147		37	37
Kalimantan Tengah	36	36	36	36	36	36	36	36	36	36	146		36	36
Kalimantan Timur	39	39	39	39	39	39	39	39	39	39	150		39	39
Kalimantan Utara	39	39	39	39	39	39	39	39	39	39	298		39	40
epulauan Bangka Belitung	38	38	38	38	38	38	38	38	38	38	150		38	38
Kepulauan Riau	39	39	39	39	39	39	39	39	39	39	149		39	39
Lampung	37	37	37	37	37	37	37	37	37	37	147		37	37
Maluku	37	37	37	37	37	37	37	37	37	37	147		37	37
Maluku Utara	36	36	36	36	36	36	36	36	36	36	344		36	36
Nusa Tenggara Barat	37	37	37	37	37	37	37	37	37	37	147		37	37
Nusa Tenggara Timur	34	34	34	34	34	34	34	34	34	34	147		34	34
Papua	37	37	37	37	37	37	37	37	37	37	197		37	37
Papua Barat	39	39	39	39	39	39	39	39	39	39	148		39	39
Riau	35	35	35	35	35	35	46	35	35	35	145		35	35
Sulawesi Barat	39	39	39	39	39	39	39	39	40	39	149		39	39
Sulawesi Selatan	34	34	34	34	34	34	34	34	34	34	144		34	34
Sulawesi Tengah	34	34	34	34	34	34	34	34	34	34	145		34	34
Sulawesi Tenggara	36	36	36	36	36	36	36	36	36	36	149		37	36
Sulawesi Utara	34	34	34	34	34	34	34	34	35	34	144		34	34
Sumatera Barat	35	35	35	35	35	35	72	35	35	35	145		35	35
Sumatera Selatan	38	38	38	38	38	38	38	38	38	38	148		38	38
Sumatera Utara	36	36	36	36	36	36	37	36	36	36	146		36	36

Lampiran 1. Tabel Jumlah Missing Value tiap Produk dan Provinsi



Lampiran 2. Visualisasi Missing Value pada Produk Cabai Rawit Merah



Lampiran 3. Visualisasi Missing Value pada tiap Produk secara Agregat