

Clustering Provinsi berdasarkan Volatilitas Harga Cabai Rawit menggunakan Fitur Volatilitas Kondisional GARCH

Provincial Clustering using GARCH-based Chili Price Volatility Features

¹Yogata Rama Guninta, ²Bety Wulan Sari*, ³Yoga Pristyanto

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
^{1,2,3}Jl. Ring Road Utara, Ngringin, Condongcatu, Kec. Depok, Kabupaten Sleman, Daerah Istimewa
Yogyakarta, Indonesia

*e-mail: bety@amikom.ac.id

(received: 4 April 2026, revised: 13 July 2026, accepted: 14 July 2026)

Abstrak

Cabai rawit merupakan komoditas pangan strategis di Indonesia yang rentan terhadap fluktuasi harga antar wilayah akibat perbedaan kondisi distribusi dan rantai pasok. Fluktuasi tersebut kerap terjadi dalam rentang waktu yang singkat, sehingga analisis berbasis data bulanan atau tahunan cenderung meredam lonjakan harga jangka pendek yang justru paling berdampak terhadap daya beli masyarakat dan kebijakan stabilisasi harga. Penelitian sebelumnya umumnya mengelompokkan wilayah berdasarkan harga nominal atau harga rata-rata sehingga belum merepresentasikan dinamika perubahan harga harian secara langsung. Penelitian ini bertujuan mengelompokkan provinsi di Indonesia berdasarkan karakteristik volatilitas harga cabai rawit, dengan mengusulkan pendekatan clustering yang menggunakan fitur volatilitas kondisional hasil pemodelan GARCH(1,1) sebagai representasi dinamika harga harian, dikombinasikan dengan algoritma K-Means untuk pengelompokannya. Data harga harian cabai rawit dari 34 provinsi selama periode April 2024 hingga April 2026 diperoleh dari Pusat Informasi Harga Pangan Strategis Nasional (PIHPS). Data ditransformasikan menjadi return harian dan dimodelkan menggunakan GARCH(1,1) untuk memperoleh nilai rata-rata volatilitas per provinsi sebagai fitur clustering. Jumlah cluster ditentukan melalui metode Elbow dan Silhouette Score. Hasil evaluasi menunjukkan konfigurasi lima cluster optimal dengan nilai silhouette sebesar 0,65, yang membagi 34 provinsi ke dalam lima kategori volatilitas: sangat tinggi, tinggi, sedang, rendah, dan sangat rendah. Temuan penelitian menunjukkan bahwa provinsi dengan tingkat harga tinggi tidak selalu berada pada kelompok dengan volatilitas tertinggi, sehingga pendekatan berbasis volatilitas mampu memberikan representasi tambahan terhadap karakteristik dinamika harga dibandingkan pendekatan berbasis harga nominal. Hasil penelitian diharapkan dapat mendukung pemantauan volatilitas harga pangan antar wilayah serta menjadi referensi dalam pengembangan sistem pendukung keputusan berbasis data.

Kata kunci: *chili pepper, conditional volatility, GARCH, K-Means clustering, price volatility.*

Abstract

Bird's eye chili is a strategic food commodity in Indonesia whose prices are highly susceptible to interregional fluctuations due to differences in distribution systems and supply chain conditions. These fluctuations often occur over short time horizons, making analyses based on monthly or annual data less capable of capturing short-term price spikes that have the greatest impact on consumers' purchasing power and price stabilization policies. Previous studies have generally clustered regions based on nominal or average prices, which do not adequately represent the dynamics of daily price movements. This study aims to cluster Indonesian provinces according to the volatility characteristics of bird's eye chili prices by proposing a clustering approach that utilizes GARCH(1,1)-based conditional volatility features to represent daily price dynamics, combined with the K-Means algorithm for cluster formation. Daily bird's eye chili price data from 34 provinces covering the period from April 2024 to April 2026 were obtained from the National Strategic Food Price Information Center (PIHPS). The price data were transformed into daily returns and modeled using

<http://sistemasi.ftik.unisi.ac.id>

GARCH(1,1) to estimate the average conditional volatility of each province, which was subsequently used as the clustering feature. The optimal number of clusters was determined using the Elbow Method and the Silhouette Score. The evaluation results identified five as the optimal number of clusters, achieving a silhouette score of 0.65 and classifying the 34 provinces into five volatility categories: very high, high, moderate, low, and very low. The findings reveal that provinces with higher price levels do not necessarily belong to the highest volatility cluster, indicating that a volatility-based approach provides additional insights into price dynamics beyond those captured by nominal price-based clustering. These results can support regional food price volatility monitoring and serve as a reference for developing data-driven decision support systems for food price management.

Keywords: *chili pepper, conditional volatility, GARCH, K-Means clustering, price volatility.*

1 Pendahuluan

Cabai rawit merah merupakan salah satu komoditas pangan strategis di Indonesia yang memiliki kontribusi signifikan terhadap inflasi daerah dan ketahanan pangan nasional [1]. Fluktuasi harga cabai yang tinggi sering memengaruhi daya beli masyarakat serta menjadi perhatian dalam pengendalian inflasi pangan oleh pemerintah daerah maupun nasional [2]. Berdasarkan data Pusat Informasi Harga Pangan Strategis Nasional (PIHPS), harga cabai rawit antar provinsi menunjukkan pola perubahan yang dinamis dan tidak seragam. Perbedaan kondisi distribusi, ketergantungan pasokan, serta gangguan rantai logistik menyebabkan beberapa wilayah mengalami gejolak harga yang lebih tinggi dibandingkan wilayah lainnya [3]. Kondisi tersebut menunjukkan pentingnya analisis volatilitas harga pada tingkat provinsi untuk mendukung pengambilan keputusan dalam distribusi dan stabilisasi pangan yang bersifat regional.

Penelitian terkait analisis harga pangan sebelumnya umumnya berfokus pada peramalan harga maupun pengelompokan wilayah berdasarkan nilai harga rata-rata atau harga absolut [4]. Algoritma K-Means banyak digunakan dalam proses clustering karena memiliki efisiensi komputasi yang baik dalam mengelompokkan karakteristik data wilayah [5]. Di sisi lain, model Generalized Autoregressive Conditional Heteroskedasticity (GARCH) telah digunakan secara luas dalam analisis deret waktu untuk memodelkan volatilitas dan menangkap fenomena volatility clustering pada data ekonomi dan keuangan [6]. Pemilihan model GARCH dalam penelitian ini didasarkan pada karakteristik data harga pangan harian yang bersifat heteroskedastis, yaitu memiliki ragam (variance) yang tidak konstan dari waktu ke waktu. Pada data semacam ini, periode gejolak harga cenderung diikuti oleh periode gejolak berikutnya (volatility clustering), sehingga pendekatan yang mengasumsikan volatilitas konstan seperti standar deviasi sederhana atau rata-rata bergerak tidak mampu menangkapnya secara memadai [7]. Berbeda dengan pendekatan tersebut, GARCH secara eksplisit memodelkan volatilitas bersyarat (conditional volatility) yang berubah mengikuti waktu, sehingga menghasilkan representasi tingkat gejolak harga yang lebih akurat dan telah terbukti efektif pada berbagai studi volatilitas harga komoditas pangan [8]. Model tersebut mampu merepresentasikan perubahan volatilitas yang bersifat dinamis pada data dengan karakteristik volatilitas yang tidak konstan, termasuk data harga pangan harian.

Meskipun demikian, sebagian besar penelitian sebelumnya masih menggunakan nilai harga sebagai fitur utama dalam proses clustering wilayah sehingga belum mampu merepresentasikan tingkat volatilitas harga secara langsung [9]. Wilayah dengan harga tinggi belum tentu memiliki volatilitas tinggi, sedangkan wilayah dengan harga relatif rendah dapat mengalami perubahan harga harian yang lebih ekstrem. Selain itu, integrasi hasil ekstraksi volatilitas ekonometrika sebagai fitur clustering wilayah masih relatif terbatas pada penelitian komoditas pangan di Indonesia [10]. Oleh karena itu, diperlukan pendekatan yang mampu memetakan karakteristik volatilitas harga berdasarkan dinamika perubahan harga, bukan hanya berdasarkan nominal harga.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan clustering provinsi berdasarkan volatilitas harga cabai rawit menggunakan fitur volatilitas kondisional GARCH. Data harga harian cabai rawit dari 34 provinsi ditransformasikan menjadi return harian (daily return) dan dimodelkan menggunakan GARCH(1,1) untuk memperoleh nilai volatilitas kondisional sebagai representasi volatilitas harga. Nilai tersebut selanjutnya digunakan sebagai fitur utama dalam proses clustering menggunakan algoritma K-Means untuk mengidentifikasi kelompok provinsi berdasarkan

<http://sistemasi.ftik.unisi.ac.id>

tingkat volatilitas harga. Berbeda dengan penelitian sebelumnya yang menggunakan harga absolut sebagai dasar clustering, penelitian ini memanfaatkan fitur volatilitas kondisional hasil pemodelan GARCH sebagai representasi dinamika harga dalam proses pengelompokan wilayah.

Penelitian ini bertujuan untuk mengidentifikasi kelompok provinsi berdasarkan karakteristik volatilitas harga cabai rawit melalui integrasi model GARCH dan clustering K-Means. Secara lebih rinci, penelitian dilakukan melalui serangkaian tahapan yang saling berurutan. Harga harian terlebih dahulu ditransformasikan menjadi return harian, kemudian fitur volatilitas kondisional diekstraksi melalui pemodelan GARCH(1,1) pada setiap provinsi. Selanjutnya, jumlah cluster optimal ditentukan menggunakan metode Elbow dan Silhouette Score, sebelum akhirnya provinsi dikelompokkan menggunakan algoritma K-Means dan diinterpretasikan karakteristik masing-masing cluster. Kontribusi utama (novelty) penelitian ini terletak pada pemanfaatan volatilitas kondisional hasil pemodelan GARCH sebagai fitur clustering, yang menjadikan GARCH bukan sebagai alat peramalan sebagaimana lazim digunakan, melainkan sebagai metode ekstraksi fitur (feature extraction) untuk pengelompokan wilayah. Hasil penelitian diharapkan dapat memberikan representasi volatilitas harga yang lebih objektif dibandingkan pendekatan berbasis harga nominal serta mendukung pengembangan sistem pendukung keputusan dalam pemantauan distribusi pangan dan mitigasi gejala harga antar wilayah.

2 Tinjauan Literatur

Penelitian terkait volatilitas harga pangan telah berkembang dalam beberapa tahun terakhir, terutama pada komoditas strategis yang memiliki pengaruh terhadap inflasi dan kestabilan pasokan pangan [1]. Fluktuasi harga pangan yang bersifat dinamis menyebabkan analisis berbasis deret waktu menjadi penting untuk memahami karakteristik perubahan harga antar periode [11]. Dalam penelitian ekonometrika, model Generalized Autoregressive Conditional Heteroskedasticity (GARCH) banyak digunakan untuk memodelkan volatilitas karena mampu menangkap fenomena volatility clustering dan volatilitas yang tidak konstan pada data ekonomi maupun keuangan [6]. Beberapa penelitian menunjukkan bahwa model GARCH efektif digunakan untuk menganalisis volatilitas harga komoditas pertanian dan pangan yang memiliki pola perubahan volatilitas tidak konstan [7]. Pendekatan tersebut dinilai lebih mampu merepresentasikan karakteristik volatilitas harga dibandingkan analisis berbasis rata-rata harga semata [8].

Selain analisis volatilitas, penelitian terkait pengelompokan wilayah berdasarkan karakteristik harga pangan juga telah banyak dilakukan menggunakan pendekatan data mining dan machine learning [12]. Algoritma K-Means menjadi salah satu metode clustering yang umum digunakan karena memiliki efisiensi komputasi dan kemampuan segmentasi data yang baik pada berbagai penelitian pengelompokan wilayah [5]. Pada penelitian sebelumnya, clustering harga pangan umumnya dilakukan menggunakan fitur harga rata-rata, tingkat inflasi, maupun statistik deskriptif sederhana sebagai dasar pembentukan kelompok wilayah [13]. Pendekatan tersebut mampu memberikan gambaran karakteristik wilayah berdasarkan kesamaan nilai data, namun belum sepenuhnya merepresentasikan tingkat volatilitas harga yang bersifat dinamis antar waktu [1].

Beberapa penelitian terkini mulai menggabungkan pendekatan ekonometrika dan machine learning untuk meningkatkan representasi karakteristik data deret waktu [14]. Meskipun demikian, integrasi hasil pemodelan volatilitas GARCH sebagai fitur utama dalam proses clustering wilayah masih relatif terbatas, khususnya pada penelitian komoditas pangan di Indonesia [11]. Sebagian besar penelitian sebelumnya masih menggunakan harga absolut atau nilai rata-rata sebagai representasi utama dalam pengelompokan wilayah sehingga perubahan volatilitas akibat fluktuasi harian belum tergambarkan secara optimal [10]. Kondisi tersebut menyebabkan wilayah dengan harga tinggi cenderung diasumsikan memiliki tingkat volatilitas yang tinggi, meskipun pada kenyataannya tingkat volatilitas antar wilayah dapat menunjukkan karakteristik yang berbeda [9].

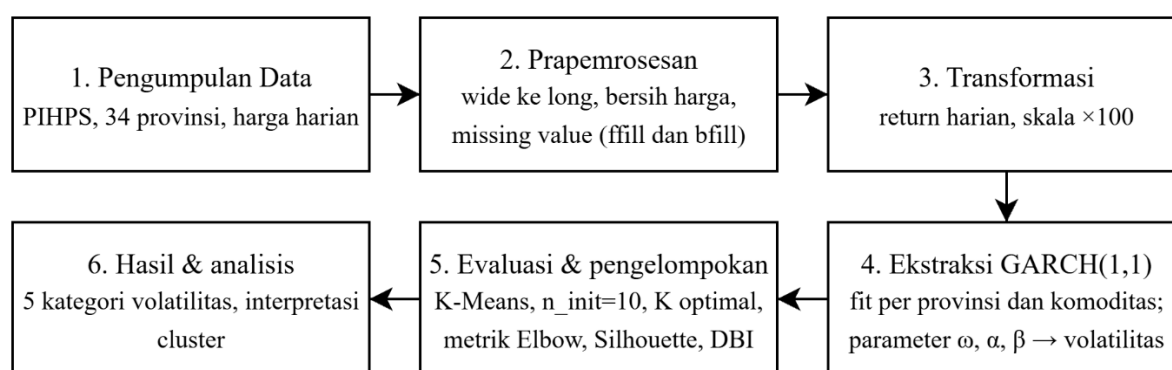
Berdasarkan analisis literatur tersebut, dapat disimpulkan bahwa penelitian sebelumnya masih memiliki keterbatasan dalam merepresentasikan karakteristik volatilitas harga sebagai dasar pengelompokan wilayah. Oleh karena itu, penelitian ini berfokus pada penggunaan fitur volatilitas kondisional hasil pemodelan GARCH sebagai representasi volatilitas harga dalam proses clustering provinsi menggunakan algoritma K-Means. Pendekatan ini diharapkan mampu memberikan pemetaan

karakteristik volatilitas harga cabai rawit yang lebih representatif dibandingkan pendekatan berbasis harga nominal atau statistik harga konvensional [5].

3 Metode Penelitian

Metode penelitian ini disusun secara sistematis menggunakan pendekatan analisis data runtun waktu (time series) untuk memetakan karakteristik volatilitas harga cabai rawit antar provinsi di Indonesia. Penelitian dilakukan menggunakan bahasa pemrograman Python versi 3.12.13 dengan memanfaatkan pustaka pandas untuk pengolahan data, scikit-learn untuk proses clustering dan evaluasi model, serta arch untuk pemodelan volatilitas GARCH [15]. Data penelitian berupa harga harian cabai rawit dari 34 provinsi yang diperoleh dari Pusat Informasi Harga Pangan Strategis Nasional (PIHPS). Pemanfaatan pustaka scikit-learn dilakukan untuk implementasi algoritma K-Means serta proses evaluasi kualitas cluster, sedangkan pustaka arch digunakan dalam proses estimasi model GARCH untuk memperoleh fitur volatilitas kondisional.

Secara umum, tahapan penelitian terdiri atas proses pengumpulan data, prapemrosesan data, transformasi data, pemodelan dan ekstraksi fitur volatilitas, evaluasi clustering, serta analisis hasil penelitian. Alur tahapan penelitian yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1 Diagram alur penelitian

Gambar 1 menunjukkan alur penelitian yang terdiri atas enam tahap. Tahap pertama adalah pengumpulan data harga harian cabai rawit dari 34 provinsi melalui PIHPS Nasional. Data kemudian melalui tahap prapemrosesan yang mencakup transformasi struktur data, pembersihan harga, serta penanganan missing value menggunakan forward fill dan backward fill. Pada tahap ketiga, harga ditransformasikan menjadi return harian sebagai dasar analisis volatilitas. Tahap keempat melakukan ekstraksi fitur volatilitas melalui pemodelan GARCH(1,1) pada setiap provinsi, menghasilkan nilai rata-rata volatilitas kondisional. Selanjutnya, tahap evaluasi dan pengelompokan menggunakan algoritma K-Means, dengan jumlah cluster optimal ditentukan melalui tiga metrik, yaitu metode Elbow, Silhouette Score, dan Davies-Bouldin Index. Tahap terakhir adalah analisis hasil, di mana provinsi dikelompokkan ke dalam lima kategori volatilitas dan diinterpretasikan karakteristiknya. Rangkaian tahap ini dirancang agar proses ekstraksi fitur volatilitas dan pengelompokan wilayah dapat dilakukan secara sistematis dan dapat direplikasi.

3.1 Pengumpulan dan Prapemrosesan Data

Data yang digunakan dalam penelitian ini merupakan data harga harian cabai rawit dari 34 provinsi di Indonesia yang bersumber dari Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) selama periode April 2024 hingga April 2026. Pemilihan unit analisis pada tingkat provinsi didasarkan pada ketersediaan dan cakupan data resmi PIHPS, yang melakukan pemantauan serta publikasi harga pangan strategis secara harian pada resolusi provinsi. Meskipun kondisi geografis, hasil alam, dan kualitas komoditas pada tingkat kabupaten atau kota dapat bervariasi dan memengaruhi pembentukan harga [3], data harga harian pada resolusi tersebut belum tersedia secara konsisten dan lengkap untuk seluruh wilayah nasional sepanjang periode pengamatan. Penggunaan tingkat provinsi juga sejalan dengan tujuan penelitian, yaitu menghasilkan pemetaan karakteristik volatilitas yang dapat mendukung kebijakan stabilisasi harga pangan yang bersifat regional [1].

<http://sistemasi.ftik.unisi.ac.id>

Keterbatasan resolusi ini selanjutnya dibahas sebagai salah satu batasan penelitian pada bagian kesimpulan.

Data harga cabai rawit yang dikumpulkan terdiri atas dua jenis komoditas, yaitu Cabai Rawit (ditandai sebagai level 1) dan Cabai Rawit Merah (ditandai sebagai level 2). Dari 34 provinsi yang diamati, sebanyak 32 provinsi memiliki data untuk kedua jenis komoditas tersebut, sedangkan dua provinsi, yaitu Aceh dan Sumatera Barat, hanya memiliki data Cabai Rawit (level 1). Perbedaan ketersediaan komoditas ini mengikuti cakupan pemantauan PIHPS pada masing-masing provinsi. Dalam penelitian ini, seluruh komoditas yang tersedia pada setiap provinsi turut diproses, sehingga nilai volatilitas suatu provinsi merepresentasikan rata-rata volatilitas dari seluruh jenis cabai rawit yang dipantau di provinsi tersebut.

Data mentah yang diperoleh dari PIHPS tersimpan dalam format lebar (wide format), di mana setiap baris merepresentasikan satu jenis komoditas dan setiap kolom merepresentasikan tanggal pengamatan. Pada tahap prapemrosesan, struktur data tersebut ditransformasikan menjadi format panjang (long format) melalui proses melting, sehingga setiap baris merepresentasikan satu observasi harga pada kombinasi provinsi, komoditas, dan tanggal tertentu. Contoh struktur data setelah transformasi ditunjukkan pada Tabel 1.

Tabel 1 Struktur data harga harian cabai rawit setelah transformasi ke format long

Provinsi	Komoditas	Level	Tanggal	Harga (Rp/kg)
Aceh	Cabai Rawit	1	30-04-2024	47.150
Aceh	Cabai Rawit	1	01-05-2024	48.500
Bali	Cabai Rawit	1	30-04-2024	35.500
Bali	Cabai Rawit Merah	2	30-04-2024	31.400
Jawa Barat	Cabai Rawit	1	30-04-2024	43.600
Jawa Barat	Cabai Rawit Merah	2	30-04-2024	48.100

Nilai harga yang semula berupa teks dengan pemisah ribuan dikonversi menjadi tipe numerik, dan entri yang tidak valid ditandai sebagai missing value. Untuk mengevaluasi kualitas data, persentase missing value dihitung untuk setiap provinsi sebelum dilakukan penanganan. Penanganan missing value dilakukan menggunakan metode forward fill dan backward fill secara berurutan pada masing-masing kombinasi provinsi dan komoditas, sehingga kekosongan nilai diisi berdasarkan kontinuitas observasi terdekat tanpa menghilangkan struktur runtun waktu. Sebagai indikator kualitas data, persentase missing value dievaluasi menggunakan ambang referensi sebesar 5% dari total observasi. Ringkasan persentase missing value sebelum dan sesudah penanganan ditunjukkan pada Tabel 2.

Tabel 2 Persentase Missing Value Sebelum dan Sesudah Penanganan

Provinsi	Missing Sebelum (%)	Missing Sesudah(%)
Kepulauan Riau	43,12	0,00
Maluku Utara	2,77	0,00
Kalimantan Barat	2,20	0,00
Sumatera Barat	2,10	0,00
Kalimantan Utara	1,63	0,00
Aceh	0,00	0,00
...	...	...
Rata-rata	1,81	0,00

Hasil evaluasi menunjukkan bahwa mayoritas provinsi memiliki persentase missing value yang rendah, dengan rata-rata sebesar 1,81% sebelum penanganan. Hanya satu provinsi, yaitu Kepulauan Riau, yang memiliki persentase missing value sebesar 43,12% dan melampaui ambang batas 5%. Untuk mengevaluasi pengaruh provinsi tersebut terhadap hasil pengelompokan, dilakukan analisis sensitivitas dengan membandingkan hasil clustering pada dua skenario, yaitu dengan menyertakan seluruh 34 provinsi dan dengan mengeksklusi Kepulauan Riau (33 provinsi). Hasil analisis sensitivitas menunjukkan bahwa eksklusi Kepulauan Riau tidak mengubah komposisi

keanggotaan cluster, di mana tidak terdapat satu pun provinsi yang berpindah cluster, dan hanya berdampak minimal pada nilai Silhouette Score, yaitu dari 0,65 menjadi 0,668. Selain tidak mengubah keanggotaan cluster, perbedaan nilai Silhouette Score yang relatif kecil menunjukkan bahwa keberadaan Provinsi Kepulauan Riau tidak memberikan pengaruh yang signifikan terhadap struktur hasil clustering. Berdasarkan hasil tersebut, penelitian ini mempertahankan seluruh 34 provinsi sebagai data utama untuk menjaga cakupan nasional yang menyeluruh, dengan seluruh missing value berhasil diimputasi menggunakan kombinasi forward fill dan backward fill sehingga tidak terdapat nilai kosong pada dataset akhir.

3.2 Transformasi Data (Perhitungan Return Harian)

Sebelum dimodelkan, data harga harian ditransformasikan menjadi return harian untuk merepresentasikan perubahan harga relatif antar hari serta mengurangi pengaruh tren dan perbedaan skala nominal harga. Return harian dihitung sebagai perubahan relatif harga pada hari ke- t terhadap harga pada hari sebelumnya, sebagaimana dinyatakan pada Persamaan (1):

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

Dengan r_t adalah return pada hari ke- t , P_t adalah harga pada hari ke- t , dan P_{t-1} adalah harga pada hari sebelumnya. Penggunaan return, dan bukan harga absolut, bertujuan agar analisis volatilitas berfokus pada dinamika perubahan harga, sehingga karakteristik gejolak harga dapat dibandingkan secara adil antar provinsi yang memiliki tingkat harga nominal yang berbeda. Selain itu, return harian merupakan bentuk data yang umum digunakan dalam pemodelan volatilitas menggunakan GARCH karena mampu merepresentasikan perubahan harga dari waktu ke waktu.

Sebagai ilustrasi teknis, Tabel 3 menampilkan contoh perhitungan return harian pada data harga cabai rawit di Provinsi Aceh. Sebagai contoh, harga pada tanggal 29 April 2024 sebesar Rp46.750 dan pada tanggal 30 April 2024 sebesar Rp47.150, sehingga return harian pada tanggal 30 April 2024 dihitung sebagai berikut:

$$r_t = \frac{47.150 - 46.750}{46.750} = 0,008556 \text{ (0,86\%)}$$

Tabel 3 Contoh transformasi harga harian menjadi return harian pada provinsi aceh

Tanggal	Harga (Rp/kg)	Perhitungan	Return
29-04-2024	46.750	-	-
30-04-2024	47.150	$(47.150 - 46.750) / 46.750$	0,008556
01-05-2024	48.500	$(48.500 - 47.150) / 47.150$	0,028632
02-05-2024	47.600	$(47.600 - 48.500) / 48.500$	-0,018557
03-05-2024	49.150	$(49.150 - 47.600) / 47.600$	0,032563

Nilai return pada tanggal pertama pengamatan tidak terdefinisi karena tidak memiliki harga hari sebelumnya, sehingga baris tersebut dikeluarkan dari analisis. Selanjutnya, nilai return dikalikan dengan konstanta 100 untuk menghasilkan return dalam satuan persen. Penskalaan ini bertujuan menjaga kestabilan numerik pada proses estimasi model GARCH, karena nilai return yang terlalu kecil dapat menyulitkan proses optimasi parameter model [15].

3.3 Pemodelan dan Ekstraksi Volatilitas GARCH(1,1)

Setelah data return harian diperoleh, tahap selanjutnya adalah mengekstraksi fitur volatilitas menggunakan model *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH). Model GARCH(1,1) dipilih karena mampu menangkap fenomena *volatility clustering*, yaitu kondisi di mana periode gejolak harga yang tinggi cenderung diikuti oleh periode gejolak yang tinggi pula [6]. Model GARCH(1,1) menyatakan bahwa varians bersyarat (*conditional variance*) pada hari ke- t , dinotasikan sebagai σ_t^2 , merupakan fungsi dari konstanta, kuadrat guncangan pada hari sebelumnya, dan varians bersyarat pada hari sebelumnya, sebagaimana dinyatakan pada Persamaan (2):

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (2)$$

Dengan σ_t^2 adalah varians bersyarat pada hari ke- t , ω adalah konstanta, ϵ_{t-1}^2 adalah kuadrat residual (guncangan) pada hari sebelumnya, σ_{t-1}^2 adalah varians bersyarat pada hari sebelumnya, serta α dan β adalah parameter model. Parameter α merepresentasikan pengaruh guncangan jangka pendek terhadap volatilitas, sedangkan parameter β merepresentasikan tingkat persistensi volatilitas. Jumlah $\alpha + \beta$ yang mendekati satu menunjukkan bahwa volatilitas bersifat persisten, yaitu gejolak harga cenderung bertahan dalam jangka waktu yang lebih lama. Parameter model diestimasi menggunakan metode Maximum Likelihood Estimation (MLE) melalui pustaka arch pada Python sehingga nilai ω , α dan β diperoleh secara numerik. Model GARCH(1,1) di-fit secara terpisah pada setiap kombinasi provinsi dan komoditas menggunakan spesifikasi mean model konstan dan distribusi normal. Spesifikasi parameter model yang digunakan dalam penelitian ini ditunjukkan pada Tabel 4.

Tabel 4 Spesifikasi parameter pemodelan GARCH(1,1)

Parameter	Nilai / Spesifikasi
Ordo model	GARCH(1,1), yaitu p = 1 dan q = 1
Mean model	Konstan (Constant)
Distribusi	Normal
Input	Return harian $\times 100$ (dalam persen)
Estimasi	Maximum Likelihood Estimation (MLE)
Pustaka	arch (Python)

Dari hasil estimasi, diperoleh nilai parameter ω , α , dan β untuk setiap kombinasi provinsi dan komoditas. Nilai volatilitas suatu provinsi kemudian dihitung sebagai rata-rata volatilitas kondisional (conditional volatility) dari seluruh komoditas yang tersedia pada provinsi tersebut. Karena input model berupa return dalam satuan persen (return $\times 100$), maka nilai rata-rata volatilitas kondisional yang dihasilkan juga dinyatakan dalam satuan persen (%), yang merepresentasikan besarnya simpangan baku bersyarat perubahan harga harian.

Sebagai ilustrasi, hasil estimasi GARCH(1,1) pada data cabai rawit di Provinsi Aceh menghasilkan parameter $\omega = 6,097$, $\alpha = 0,321$, dan $\beta = 0,445$, dengan nilai rata-rata volatilitas kondisional sebesar 4,433%. Nilai $\alpha + \beta = 0,766$ menunjukkan adanya persistensi volatilitas pada harga cabai rawit di provinsi tersebut. Contoh nilai parameter dan volatilitas kondisional untuk beberapa provinsi ditunjukkan pada Tabel 5.

Tabel 5 Parameter GARCH(1,1) dan rata-rata volatilitas kondisional per provinsi

Provinsi	ω	α	β	$\alpha + \beta$	Rata-rata Volatilitas Kondisional (%)
Gorontalo	2,487	0,028	0,927	0,955	7,311
Papua Barat	26,689	0,152	0,357	0,508	7,213
Sulawesi Utara	8,226	0,000	0,748	0,748	6,748
Nusa Tenggara Barat	10,178	0,019	0,757	0,777	6,671
Bali	7,601	0,053	0,796	0,849	6,410
Sulawesi Tengah	12,991	0,054	0,573	0,626	6,063
Maluku	1,029	0,012	0,960	0,972	5,886
Banten	14,732	0,113	0,435	0,548	5,884
Kalimantan Timur	9,719	0,067	0,682	0,748	5,881
Sumatera Utara	4,001	0,295	0,647	0,942	5,814

Provinsi	ω	α	β	$\alpha + \beta$	Rata-rata Volatilitas Kondisional (%)
Jambi	28,901	0,198	0,000	0,198	5,798
Jawa Timur	28,524	0,072	0,059	0,132	5,688
DI Yogyakarta	5,236	0,003	0,835	0,838	5,516
Kalimantan Selatan	0,742	0,013	0,961	0,974	5,190
Jawa Tengah	19,186	0,210	0,115	0,325	5,151
DKI Jakarta	2,710	0,050	0,859	0,909	5,141
Sulawesi Selatan	0,100	0,000	0,998	0,998	5,103
Kalimantan Barat	23,999	0,096	0,000	0,096	5,051
Kepulauan Bangka Belitung	0,161	0,000	0,996	0,996	4,942
Kalimantan Utara	4,762	0,099	0,726	0,825	4,869
Sumatera Selatan	6,128	0,047	0,676	0,723	4,602
Aceh	6,097	0,321	0,445	0,766	4,433
Lampung	16,467	0,206	0,000	0,206	4,428
Maluku Utara	16,517	0,185	0,000	0,185	4,333
Bengkulu	11,631	0,428	0,022	0,450	4,287
Sulawesi Barat	8,674	0,011	0,469	0,480	4,272
Riau	8,601	0,430	0,363	0,793	4,256
Kalimantan Tengah	7,403	0,040	0,533	0,573	4,184
Sumatera Barat	0,304	0,026	0,960	0,985	4,161
Kepulauan Riau	22,748	0,110	0,407	0,517	3,889
Papua	5,683	0,066	0,518	0,583	3,742
Jawa Barat	0,974	0,102	0,844	0,946	3,705
Sulawesi Tenggara	0,315	0,061	0,923	0,984	3,573
Nusa Tenggara Timur	3,217	0,053	0,582	0,635	2,750

Keterangan: Nilai $\alpha + \beta$ dihitung dari parameter penuh sebelum pembulatan, sehingga dapat berbeda $\pm 0,001$ dari hasil penjumlahan nilai yang ditampilkan pada tabel.

Berdasarkan Tabel 5, terlihat bahwa nilai parameter GARCH bervariasi antar provinsi, yang mencerminkan perbedaan karakteristik dinamika harga di masing-masing wilayah. Sebagian besar provinsi menunjukkan nilai β yang lebih besar dibandingkan α , yang mengindikasikan bahwa volatilitas harga lebih dipengaruhi oleh persistensi gejolak dari periode sebelumnya dibandingkan guncangan sesaat. Sebagai contoh, Provinsi Gorontalo memiliki nilai $\beta = 0,927$ dengan $\alpha + \beta = 0,955$ yang mendekati satu, menunjukkan bahwa gejolak harga di wilayah tersebut cenderung bertahan lama. Perlu dicatat pula bahwa nilai rata-rata volatilitas kondisional yang tinggi tidak selalu sejalan dengan tingkat harga nominal maupun besarnya parameter tertentu. Sebagai contoh, Papua Barat memiliki nilai ω yang jauh lebih besar dibandingkan provinsi lain, namun rata-rata volatilitas kondisionalnya berada pada tingkat yang serupa dengan Gorontalo. Hal ini menunjukkan bahwa volatilitas harga merupakan hasil interaksi dari ketiga parameter model secara bersamaan, sehingga penggunaan rata-rata volatilitas kondisional sebagai fitur pengelompokan mampu merepresentasikan karakteristik dinamika harga secara lebih menyeluruh dibandingkan pendekatan berbasis nilai tunggal.

3.4 Evaluasi dan Analisis Clustering K-Means

Nilai rata-rata volatilitas kondisional yang diperoleh dari tahap sebelumnya digunakan sebagai fitur tunggal dalam proses pengelompokan provinsi menggunakan algoritma K-Means.

Algoritma K-Means mengelompokkan data dengan meminimalkan jarak antara setiap titik data terhadap pusat cluster (centroid) terdekat, sebagaimana dinyatakan pada Persamaan (3):

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3)$$

Dengan J adalah total jarak kuadrat dalam cluster (inertia), k adalah jumlah cluster, C_i adalah cluster ke- i , x adalah titik data (nilai volatilitas provinsi), dan μ_i adalah centroid cluster ke- i .

Dalam penelitian ini, algoritma K-Means dijalankan dengan parameter jumlah inisialisasi ulang (n_init) sebesar 10, yang berarti algoritma dijalankan sebanyak sepuluh kali dengan inisialisasi centroid yang berbeda, kemudian dipilih hasil dengan nilai inertia terkecil. Penggunaan n_init sebesar 10 bertujuan untuk mengurangi risiko konvergensi pada solusi local optima yang dapat terjadi akibat inisialisasi centroid awal yang kurang tepat [5]. Selain itu, parameter $random_state$ diatur bernilai 42 untuk menjaga replikabilitas hasil pengelompokan.

Penentuan jumlah cluster optimal dievaluasi menggunakan tiga metrik, yaitu metode *Elbow* berdasarkan nilai *inertia*, *Silhouette Score*, dan *Davies-Bouldin Index* (DBI). Metode *Elbow* mengamati titik di mana penurunan *inertia* mulai melandai terhadap penambahan jumlah cluster. *Silhouette Score* mengukur seberapa baik suatu data berada dalam cluster-nya dibandingkan cluster lain, dengan rentang nilai dari -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan pemisahan cluster yang lebih baik. *Davies-Bouldin Index* mengukur rasio antara sebaran di dalam cluster dan jarak antar cluster, di mana nilai yang lebih rendah menunjukkan kualitas cluster yang lebih baik [16].

Silhouette Score dihitung berdasarkan Persamaan (4), di mana untuk setiap titik data dihitung selisih antara rata-rata jarak terhadap titik lain dalam cluster yang sama (a) dan rata-rata jarak terhadap titik pada cluster terdekat (b).

$$s = \frac{b - a}{\max(a, b)} \quad (4)$$

Dengan s adalah nilai *silhouette* suatu titik data, a adalah rata-rata jarak intra-cluster, dan b adalah rata-rata jarak terhadap cluster terdekat. Nilai *Silhouette Score* akhir merupakan rata-rata nilai s dari seluruh titik data. Sebagai ilustrasi, jika sebuah provinsi memiliki rata-rata jarak intra-cluster $a = 0,15$ dan rata-rata jarak terhadap cluster terdekat $b = 0,60$, maka nilai *silhouette* provinsi tersebut adalah $s = (0,60 - 0,15) / \max(0,15; 0,60) = 0,75$, yang menunjukkan bahwa provinsi tersebut terkelompok dengan baik.

Evaluasi dilakukan untuk rentang jumlah cluster $k = 2$ hingga $k = 10$. Hasil perhitungan ketiga metrik untuk setiap nilai k ditunjukkan pada Tabel 6, yang selanjutnya menjadi dasar penentuan jumlah cluster optimal pada bagian hasil dan pembahasan.

Tabel 6 Hasil evaluasi clustering untuk berbagai jumlah cluster (K)

K	INERTIA	SILHOUETTE SCORE	DAVIES-BOULDIN INDEX
2	12,131	0,595	0,534
3	5,969	0,588	0,510
4	3,706	0,594	0,457
5	1,814	0,650	0,427
6	1,050	0,668	0,304
7	0,539	0,691	0,274
8	0,417	0,647	0,331
9	0,303	0,611	0,370
10	0,252	0,593	0,319

Berdasarkan Tabel 6, nilai *inertia* menurun secara konsisten seiring bertambahnya jumlah cluster, sebagaimana diharapkan pada metode *Elbow*, dengan penurunan tajam yang mulai melandai

<http://sistemasi.ftik.unisi.ac.id>

di sekitar $k = 5$. Sementara itu, nilai *Silhouette Score* dan *Davies-Bouldin Index* menunjukkan variasi pada rentang k yang diuji, di mana beberapa nilai k memberikan kualitas metrik yang berdekatan. Oleh karena itu, penentuan jumlah cluster optimal tidak hanya didasarkan pada satu metrik tunggal, melainkan mempertimbangkan kombinasi ketiga metrik tersebut bersama dengan aspek interpretabilitas hasil. Analisis lebih lanjut mengenai pemilihan jumlah cluster optimal beserta justifikasinya diuraikan pada bagian hasil dan pembahasan.

4 Hasil dan Pembahasan

Bagian ini menyajikan hasil pengolahan dan analisis data berdasarkan tahapan penelitian yang telah dilakukan. Pembahasan dimulai dari hasil transformasi data dan analisis statistik deskriptif return, dilanjutkan dengan pemodelan volatilitas menggunakan GARCH(1,1), evaluasi pembentukan cluster, serta analisis hasil pengelompokan provinsi berdasarkan karakteristik volatilitas harga cabai rawit. Setiap tahapan dianalisis untuk mengevaluasi kemampuan model dalam merepresentasikan perbedaan tingkat volatilitas harga antar wilayah.

4.1 Hasil Prapemrosesan dan Transformasi Data

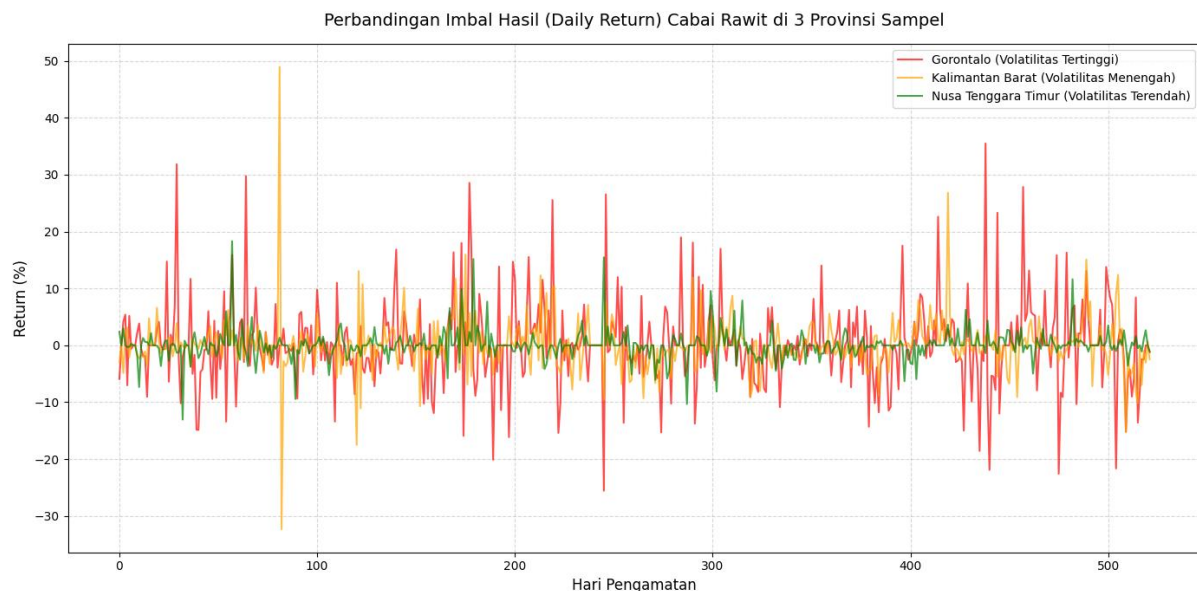
Tahap awal penelitian dilakukan melalui transformasi data harga harian cabai rawit menjadi nilai return harian. Transformasi ini dilakukan untuk merepresentasikan perubahan harga relatif antar periode serta membantu mengurangi pengaruh tren pada data runtun waktu sebelum dilakukan pemodelan volatilitas. Untuk memperoleh gambaran awal karakteristik return harian antar provinsi, dilakukan analisis statistik deskriptif yang mencakup nilai rata-rata (mean), standar deviasi, skewness, dan kurtosis. Ringkasan statistik deskriptif return untuk beberapa provinsi ditunjukkan pada Tabel 7.

Tabel 7 Statistik deskriptif return harian beberapa provinsi sampel

Provinsi	Mean	Standar Deviasi	Skewness	Kurtosis
Gorontalo	0,00209	0,07324	0,752	3,721
Papua Barat	0,00197	0,07254	1,904	11,359
Bali	0,00292	0,06629	2,444	16,953
Kalimantan Barat	0,00109	0,05320	4,350	80,940
Jawa Barat	0,00122	0,03759	0,975	6,482
Nusa Tenggara Timur	0,00069	0,02772	1,144	13,258

Berdasarkan Tabel 7, seluruh provinsi menunjukkan nilai skewness yang positif dan nilai kurtosis yang jauh lebih besar dari nol. Nilai kurtosis yang tinggi mengindikasikan bahwa distribusi return harian bersifat *leptokurtic* atau berekor gemuk (*fat-tailed*), yang berarti lonjakan harga ekstrem terjadi lebih sering dibandingkan yang diperkirakan pada distribusi normal. Sementara itu, nilai skewness yang positif menunjukkan bahwa kenaikan harga yang tajam cenderung lebih dominan dibandingkan penurunan harga. Karakteristik distribusi tersebut, yang ditunjukkan oleh nilai skewness positif dan kurtosis yang tinggi, menunjukkan bahwa data return tidak mengikuti distribusi normal. Bersama dengan pola *volatility clustering* yang teramati pada data return harian, karakteristik tersebut mendukung penggunaan model GARCH untuk memodelkan volatilitas yang berubah terhadap waktu.

Selanjutnya, untuk mengilustrasikan perbedaan karakteristik volatilitas antar wilayah, dipilih tiga provinsi sebagai representasi berdasarkan tingkat volatilitasnya, yaitu provinsi dengan volatilitas tertinggi, menengah, dan terendah. Provinsi Gorontalo dipilih sebagai representasi volatilitas tertinggi, Kalimantan Barat sebagai representasi volatilitas menengah, dan Nusa Tenggara Timur sebagai representasi volatilitas terendah. Pemilihan ketiga provinsi ini bertujuan untuk memperlihatkan rentang variasi pola pergerakan return secara kontras, dari yang paling fluktuatif hingga yang paling stabil, tanpa perlu menampilkan seluruh 34 provinsi yang akan menyulitkan interpretasi visual. Visualisasi perbandingan return harian pada ketiga provinsi tersebut ditunjukkan pada Gambar 2.



Gambar 2 Perbandingan return harian cabai rawit di tiga provinsi sampel

Berdasarkan Gambar 2, terlihat bahwa pola pergerakan return harian pada masing-masing provinsi menunjukkan tingkat fluktuasi yang berbeda. Provinsi Gorontalo memiliki perubahan return yang relatif lebih tinggi dengan beberapa lonjakan nilai pada periode tertentu, sedangkan Provinsi Nusa Tenggara Timur menunjukkan pola perubahan return yang lebih stabil dan cenderung berada di sekitar garis nol. Sementara itu, Kalimantan Barat memperlihatkan pola volatilitas menengah dengan fluktuasi yang tidak setinggi Gorontalo namun lebih dinamis dibandingkan Nusa Tenggara Timur. Selain itu, pada beberapa periode terlihat adanya pengelompokan fluktuasi tinggi dan rendah secara berurutan yang mengindikasikan pola *volatility clustering* pada data return harian. Karakteristik tersebut mendukung penggunaan model GARCH dalam memodelkan perubahan volatilitas yang bersifat dinamis pada data harga cabai rawit antar provinsi. Perbedaan pola fluktuasi tersebut menunjukkan bahwa masing-masing provinsi memiliki karakteristik volatilitas yang berbeda sehingga layak dijadikan dasar dalam proses pengelompokan menggunakan K-Means.

4.2 Hasil Ekstraksi Volatilitas GARCH(1,1)

Data return harian dari 34 provinsi selanjutnya dimodelkan menggunakan GARCH(1,1) untuk memperoleh nilai volatilitas harga yang bersifat dinamis. Pemodelan dilakukan secara terpisah dan independen pada masing-masing kombinasi provinsi dan komoditas, sehingga estimasi model menghasilkan nilai volatilitas yang berbeda pada setiap wilayah sesuai dengan karakteristik perubahan harga yang dimilikinya. Nilai rata-rata volatilitas kondisional beserta parameter GARCH(1,1) untuk seluruh provinsi telah disajikan sebelumnya pada Tabel 5.

Hasil estimasi menunjukkan rentang nilai rata-rata volatilitas kondisional yang cukup lebar antar provinsi, yaitu dari sekitar 2,75% pada provinsi dengan volatilitas terendah hingga 7,31% pada provinsi dengan volatilitas tertinggi. Provinsi Gorontalo, Papua Barat, dan Sulawesi Utara berada pada tingkat volatilitas tertinggi, sedangkan Nusa Tenggara Timur, Sulawesi Tenggara, dan Jawa Barat berada pada tingkat volatilitas terendah. Rentang nilai yang lebar ini mengindikasikan bahwa karakteristik fluktuasi harga cabai rawit antar provinsi tidak bersifat homogen, sehingga terdapat dasar yang kuat untuk mengelompokkan provinsi berdasarkan tingkat volatilitasnya.

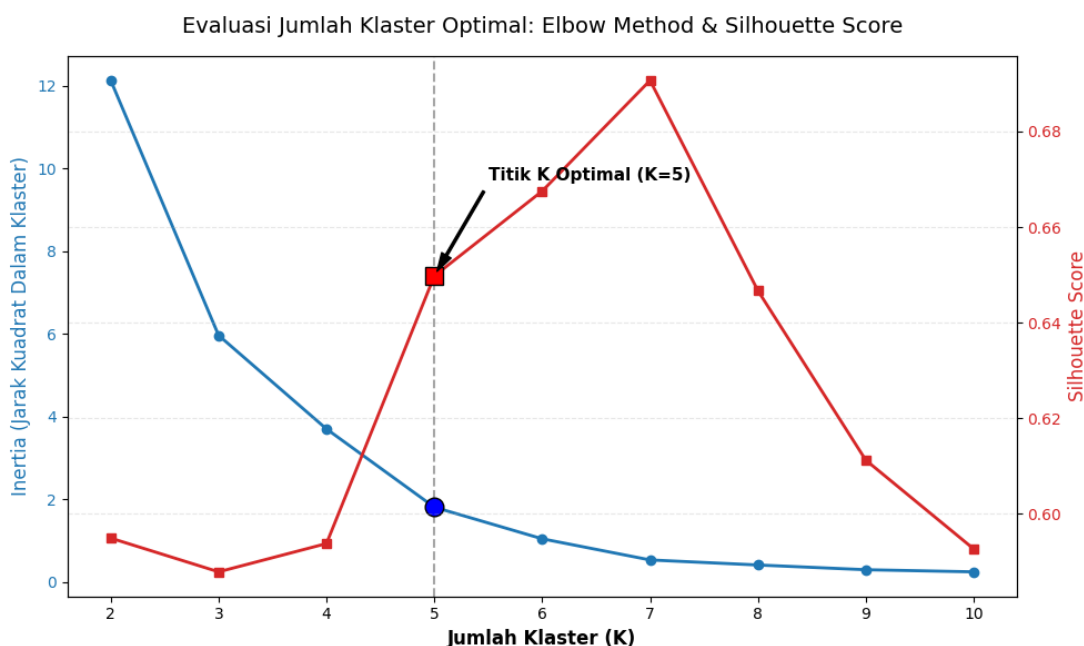
Konsistensi hasil pemodelan GARCH dapat diamati dengan membandingkannya terhadap statistik deskriptif return pada Tabel 7. Provinsi yang memiliki standar deviasi return tinggi, seperti Gorontalo dan Papua Barat, juga menempati posisi volatilitas kondisional tertinggi hasil GARCH, sedangkan provinsi dengan standar deviasi rendah, seperti Nusa Tenggara Timur, konsisten berada pada tingkat volatilitas terendah. Kesesuaian arah ini menunjukkan bahwa nilai volatilitas kondisional yang dihasilkan model tidak bertentangan dengan sebaran return yang teramati secara langsung. Meskipun demikian, berbeda dengan standar deviasi yang bersifat statis dan mengasumsikan volatilitas konstan sepanjang periode, volatilitas kondisional GARCH mampu menangkap perubahan

tingkat gejala dari waktu ke waktu, sehingga memberikan representasi dinamika harga yang lebih kaya sebagai dasar pengelompokan.

Salah satu temuan penting pada tahap ini adalah bahwa tingkat volatilitas suatu provinsi tidak selalu sejalan dengan besarnya parameter individual GARCH maupun tingkat harga nominalnya. Sebagaimana terlihat pada Tabel 5, beberapa provinsi dengan nilai konstanta (ω) yang besar tidak selalu memiliki volatilitas kondisional tertinggi, karena nilai volatilitas merupakan hasil interaksi ketiga parameter model (ω , α , dan β) secara bersamaan. Hal ini menegaskan bahwa rata-rata volatilitas kondisional merupakan fitur yang lebih representatif untuk menggambarkan dinamika harga dibandingkan penggunaan nilai tunggal seperti rata-rata harga atau standar deviasi sederhana. Nilai rata-rata volatilitas kondisional inilah yang selanjutnya digunakan sebagai fitur utama dalam proses clustering menggunakan algoritma K-Means.

4.3 Evaluasi Pembentukan Cluster

Fitur volatilitas hasil estimasi GARCH selanjutnya digunakan sebagai input utama dalam proses clustering menggunakan algoritma K-Means. Evaluasi jumlah cluster dilakukan untuk menentukan konfigurasi pengelompokan yang paling representatif terhadap karakteristik volatilitas harga antar provinsi. Evaluasi ini menggunakan tiga metrik, yaitu metode Elbow berdasarkan nilai inertia, Silhouette Score, dan Davies-Bouldin Index (DBI), sebagaimana hasil selengkapnya telah disajikan pada Tabel 6.



Gambar 3 Evaluasi jumlah cluster optimal: *elbow method* & *silhouette score*

Merujuk pada Gambar 3, kurva inertia menunjukkan penurunan yang tajam pada awal penambahan jumlah cluster dan mulai melandai membentuk siku (elbow) di sekitar titik $K = 5$. Titik ini mengindikasikan bahwa penambahan jumlah cluster setelah lima kelompok tidak lagi memberikan penurunan inertia yang berarti secara proporsional.

Berdasarkan Tabel 6, nilai Silhouette Score tertinggi justru diperoleh pada $K = 7$ sebesar 0,691, sedikit lebih tinggi dibandingkan $K = 5$ yang bernilai 0,65. Namun, selisih peningkatan tersebut hanya sebesar 0,041, yang tergolong kecil dan tidak signifikan secara praktis. Pola serupa juga teramati pada Davies-Bouldin Index, di mana nilai pada $K = 7$ (0,274) hanya sedikit lebih rendah dibandingkan $K = 5$ (0,427). Meskipun secara metrik $K = 7$ tampak sedikit lebih unggul, peningkatan yang marginal tersebut tidak sebanding dengan bertambahnya kompleksitas interpretasi akibat penambahan dua kelompok cluster.

Penentuan jumlah cluster optimal dalam penelitian ini mempertimbangkan keseimbangan antara kualitas metrik dan interpretabilitas hasil sesuai prinsip *parsimony*, yaitu memilih model yang

lebih sederhana ketika peningkatan kualitas dari model yang lebih kompleks tergolong marginal. Konfigurasi lima cluster dipilih karena beberapa pertimbangan. Pertama, nilai silhouette pada $K = 5$ sebesar 0,65 sudah menunjukkan kualitas pemisahan cluster yang baik. Kedua, titik siku pada metode Elbow berada di $K = 5$, sehingga konfigurasi ini didukung oleh dua metrik sekaligus. Ketiga, jumlah cluster yang lebih sedikit menghasilkan lima kategori volatilitas yang lebih mudah diinterpretasikan dan lebih aplikatif untuk mendukung analisis kebijakan, dibandingkan tujuh kelompok yang berisiko memecah wilayah dengan karakteristik serupa menjadi kelompok-kelompok kecil yang kurang bermakna secara praktis. Berdasarkan pertimbangan tersebut, konfigurasi lima cluster digunakan pada tahap clustering akhir untuk memetakan karakteristik volatilitas harga cabai rawit antar provinsi.

4.4 Analisis Hasil Clustering Volatilitas

Hasil clustering menggunakan algoritma K-Means berhasil mengelompokkan 34 provinsi di Indonesia ke dalam lima kelompok berdasarkan karakteristik volatilitas harga cabai rawit. Ringkasan karakteristik masing-masing cluster berdasarkan nilai rata-rata volatilitas ditunjukkan pada Tabel 8, sedangkan hasil pengelompokan provinsi pada setiap cluster ditampilkan pada Tabel 9.

Tabel 8 Karakteristik Profil Volatilitas per Cluster

Kategori Volatilitas	Jumlah Provinsi	Rata-rata Volatilitas (%)	Min	Max
Sangat Tinggi	5	6,87	6,41	7,31
Tinggi	8	5,82	5,52	6,06
Sedang	7	5,06	4,87	5,19
Rendah	9	4,33	4,16	4,60
Sangat Rendah	5	3,53	2,75	3,89

Berdasarkan Tabel 8, terlihat bahwa setiap cluster memiliki rentang nilai volatilitas yang berbeda dan tidak saling tumpang tindih, yang menunjukkan bahwa pengelompokan berhasil memisahkan provinsi berdasarkan tingkat volatilitasnya secara jelas. Kelompok Sangat Tinggi memiliki rata-rata volatilitas sebesar 6,87%, sedangkan kelompok Sangat Rendah memiliki rata-rata 3,53%, sehingga terdapat rentang yang cukup lebar antara kelompok dengan volatilitas tertinggi dan terendah.

Tabel 9 Hasil Pengelompokan 34 Provinsi Berdasarkan Fitur Volatilitas GARCH

Cluster Volatilitas	Daftar Provinsi Anggota Cluster
Sangat Tinggi	Gorontalo, Papua Barat, Sulawesi Utara, Nusa Tenggara Barat, Bali
Tinggi	Sulawesi Tengah, Maluku, Banten, Kalimantan Timur, Sumatera Utara, Jambi, Jawa Timur, DI Yogyakarta
Sedang	Kalimantan Selatan, Jawa Tengah, DKI Jakarta, Sulawesi Selatan, Kalimantan Barat, Kep. Bangka Belitung, Kalimantan Utara
Rendah	Sumatera Selatan, Aceh, Lampung, Maluku Utara, Bengkulu, Sulawesi Barat, Riau, Kalimantan Tengah, Sumatera Barat
Sangat Rendah	Kepulauan Riau, Papua, Jawa Barat, Sulawesi Tenggara, Nusa Tenggara Timur

Cluster dengan kategori volatilitas sangat tinggi mencakup lima provinsi, yaitu Gorontalo, Papua Barat, Sulawesi Utara, Nusa Tenggara Barat, dan Bali. Tingginya volatilitas pada wilayah-wilayah tersebut diduga berkaitan dengan faktor distribusi dan ketergantungan pasokan antar daerah, mengingat gangguan pada rantai pasok dapat memicu perubahan harga yang tajam pada komoditas pangan strategis [3]. Perbedaan kondisi distribusi dan disparitas harga antar wilayah di Indonesia juga telah diidentifikasi sebagai salah satu penyebab dinamika harga cabai rawit yang tidak seragam [1]. Meskipun demikian, faktor-faktor tersebut berada di luar cakupan data yang dianalisis dalam penelitian ini, sehingga penjelasan mengenai penyebab spesifik tingginya volatilitas pada masing-masing wilayah memerlukan kajian lebih lanjut. Sebaliknya, provinsi pada kategori sangat rendah, seperti Nusa Tenggara Timur dan Jawa Barat, menunjukkan pola perubahan harga yang relatif lebih stabil sepanjang periode pengamatan.

Hasil pengelompokan pada Tabel 9 juga menunjukkan bahwa provinsi dengan karakteristik volatilitas serupa dapat berada pada wilayah geografis yang berjauhan. Sebagai contoh, DI Yogyakarta yang berada di Pulau Jawa berada pada kategori volatilitas tinggi, sedangkan DKI Jakarta yang secara geografis berdekatan justru berada pada kategori sedang. Hal ini mengindikasikan bahwa pola volatilitas harga tidak semata-mata ditentukan oleh kedekatan geografis, melainkan lebih dipengaruhi oleh dinamika perubahan harga yang spesifik pada masing-masing wilayah. Selain itu, temuan penting lainnya adalah bahwa provinsi dengan tingkat harga nominal yang tinggi tidak selalu berada pada cluster dengan volatilitas tertinggi, sehingga pendekatan berbasis fitur volatilitas mampu memberikan representasi tambahan terhadap karakteristik dinamika harga yang tidak dapat ditangkap oleh pendekatan berbasis harga nominal semata.

Untuk menegaskan posisi dan kontribusi penelitian ini, hasil yang diperoleh perlu dibandingkan dengan penelitian-penelitian sebelumnya yang serupa. Secara umum, penelitian terdahulu di bidang ini dapat dikelompokkan ke dalam dua arah utama yang masih terpisah.

Kelompok pertama adalah penelitian yang melakukan pengelompokan wilayah berdasarkan karakteristik harga pangan, namun menggunakan nilai harga atau inflasi secara langsung sebagai fitur. Penelitian oleh Zulhijah dan Suarna [5] serta Sari dkk. [13] mengelompokkan wilayah berdasarkan indikator inflasi menggunakan algoritma clustering, sedangkan Tsabitah dkk. [10] melakukan spatiotemporal clustering harga komoditas pangan. Pendekatan-pendekatan tersebut mampu memetakan kesamaan tingkat harga antar wilayah, namun belum merepresentasikan dinamika volatilitas harga yang bersifat berubah terhadap waktu, karena fitur yang digunakan bersifat statis dan tidak menangkap fenomena volatility clustering.

Kelompok kedua adalah penelitian yang menerapkan model GARCH untuk menganalisis volatilitas harga pangan, namun berhenti pada tahap estimasi dan peramalan tanpa dilanjutkan ke pengelompokan wilayah. Matondang dkk. [8] menganalisis fluktuasi dan volatilitas komoditas pangan strategis di Indonesia, sementara Aravalath dan Dutta [7] serta Dinku dan Worku [6] memodelkan volatilitas harga pangan menggunakan berbagai varian GARCH. Penelitian-penelitian tersebut berhasil menangkap karakteristik volatilitas harga secara akurat, namun hasil estimasi volatilitasnya tidak dimanfaatkan lebih lanjut sebagai dasar untuk mengelompokkan wilayah secara sistematis.

Penelitian yang paling mendekati pendekatan dalam studi ini adalah penelitian yang mengombinasikan machine learning untuk pengelompokan regional harga pangan, seperti Liu dkk. [12] yang menganalisis cluster regional harga telur di Cina, serta Aguilera dan Yazdi [9] yang melakukan clustering regional harga sayur dan buah di Chile. Meskipun secara struktur serupa, kedua penelitian tersebut tidak menggunakan volatilitas kondisional hasil pemodelan GARCH sebagai fitur pengelompokan, dan dilakukan pada konteks komoditas serta wilayah yang berbeda dari Indonesia.

Berdasarkan perbandingan tersebut, terlihat adanya celah penelitian (research gap) yang menjadi fokus studi ini, yaitu belum adanya integrasi antara model GARCH sebagai metode ekstraksi fitur volatilitas dan algoritma K-Means sebagai metode pengelompokan wilayah, khususnya pada komoditas cabai rawit dalam konteks 34 provinsi di Indonesia. Penelitian ini mengisi celah tersebut dengan memanfaatkan volatilitas kondisional GARCH sebagai fitur clustering, sehingga menjadikan GARCH bukan sekadar alat peramalan, melainkan metode ekstraksi fitur untuk pengelompokan wilayah. Kontribusi ini menghasilkan pemetaan karakteristik volatilitas harga yang lebih representatif dibandingkan pendekatan berbasis harga nominal, sekaligus melengkapi arah penelitian sebelumnya yang selama ini berjalan terpisah antara analisis volatilitas dan pengelompokan wilayah. Temuan ini

<http://sistemasi.ftik.unisi.ac.id>

menunjukkan bahwa pendekatan yang diusulkan tidak hanya menggabungkan dua metode yang sebelumnya berdiri sendiri, tetapi juga menghasilkan representasi karakteristik volatilitas yang lebih sesuai untuk kebutuhan segmentasi wilayah.

5 Kesimpulan

Penelitian ini menghasilkan pengelompokan 34 provinsi di Indonesia ke dalam lima kategori volatilitas harga cabai rawit melalui integrasi model GARCH(1,1) sebagai metode ekstraksi fitur dan algoritma K-Means sebagai metode pengelompokan, dengan konfigurasi optimal pada lima cluster dan nilai Silhouette Score sebesar 0,65. Provinsi dengan volatilitas tertinggi meliputi Gorontalo, Papua Barat, Sulawesi Utara, Nusa Tenggara Barat, dan Bali dengan rata-rata 6,87%, sedangkan volatilitas terendah mencakup Nusa Tenggara Timur, Sulawesi Tenggara, dan Jawa Barat dengan rata-rata 3,53%. Temuan utama menunjukkan bahwa provinsi dengan harga nominal tinggi tidak selalu memiliki volatilitas tertinggi, sehingga pendekatan berbasis volatilitas memberikan representasi dinamika harga yang tidak tertangkap oleh pendekatan harga nominal. Secara teoretis, penelitian ini memperluas pemanfaatan GARCH dari alat peramalan menjadi metode ekstraksi fitur untuk pengelompokan wilayah. Secara praktis, hasil penelitian dapat menjadi dasar bagi pemangku kebijakan dalam memprioritaskan pemantauan harga pada wilayah bervolatilitas tinggi. Penelitian ini terbatas pada penggunaan satu fitur volatilitas, resolusi tingkat provinsi, dan belum melibatkan faktor eksternal seperti distribusi dan produksi. Oleh karena itu, penelitian selanjutnya disarankan menambahkan fitur multivariat seperti parameter persistensi ($\alpha + \beta$) atau skewness dan kurtosis return, mengintegrasikan variabel eksternal, serta membandingkan K-Means dengan metode lain seperti hierarchical clustering atau DBSCAN untuk menguji kestabilan hasil pengelompokan.

Referensi

- [1] I. R. A. S. Putri, Harianto, and N. Rosiana, "Disparitas Harga Cabai Rawit (*Capsicum frutescens* L.) Antar Waktu dan Antar Wilayah di Indonesia," *Jurnal Agribisnis Indonesia (Journal of Indonesian Agribusiness)*, Vol. 13, No. 2, pp. 333–349, 2025, DOI: 10.29244/jai.2025.13.2.333-349.
- [2] M. A. Kaban, J. R. Mandei, and M. H. Montolalu, "Pengaruh Harga Komoditas Pangan Nabati terhadap Inflasi di Kota Manado," *Agri-SosioEkonomi*, Vol. 20, No. 3, pp. 885–892, 2024.
- [3] I. Marina, D. Sukmawati, E. Juliana, and N. Safa, "Dinamika Pasar Komoditas Pangan Strategis: Analisis Fluktuasi Harga dan Produksi," *Paspalum: Jurnal Ilmiah Pertanian*, Vol. 12, No. 1, pp. 160–168, 2024, DOI: 10.35138/paspalum.v12i1.700.
- [4] T. Ulussever, H. M. Ertuğrul, S. K. Depren, M. T. Kartal, and Ö. Depren, "Estimation of Impacts of Global Factors on World Food Prices: A Comparison of Machine Learning Algorithms and Time Series Econometric Models," *Foods*, Vol. 12, No. 4, 2023, DOI: 10.3390/foods12040873.
- [5] A. Zulhijah and N. Suarna, "Penerapan Clustering Nilai Inflasi berdasarkan Kelompok Pengeluaran menggunakan Algoritma K-Means di Jawa Barat," *Jurnal Teknik Mesin, Industri, Elektro dan Informatika (JTMEI)*, Vol. 2, No. 1, pp. 81–94, 2023.
- [6] T. Dinku and G. Worku, "Asymmetric GARCH Models on Price Volatility of Agricultural Commodities," *SN Business & Economics*, Vol. 2, 2022, DOI: 10.1007/s43546-022-00355-7.
- [7] L. M. Aravalath and S. Dutta, "Forecasting World Food Price Volatility: Performance of the GARCH Model with Different Distributions Assumptions," *Economic Alternatives*, Vol. 2024, No. 1, pp. 120–141, 2024, DOI: 10.37075/EA.2024.1.08.
- [8] M. R. Matondang, B. Krisnamurthi, and Herawati, "Price Fluctuations and Volatility of National Strategic Food Commodities in Indonesia," *Agrisocionomics: Jurnal Sosial Ekonomi dan Kebijakan Pertanian*, Vol. 8, No. 1, pp. 134–146, 2024, DOI: 10.14710/agrisocionomics.v8i1.17753.
- [9] S. G. Aguilera and A. K. Yazdi, "Anomaly Detection and Regional Clustering in Chilean Wholesale Fruit and Vegetable Prices with Machine Learning," *Agriculture*, Vol. 15, No. 22, p. 2362, 2025, DOI: 10.3390/Agriculture15222362.

- [10] D. U. Tsabitah, Y. Angraini, and I. M. Sumertajaya, “Spatiotemporal Clustering of Key Food Commodity Prices using Multivariate Time Series,” *International Journal of Advances in Data and Information Systems*, Vol. 6, No. 3, pp. 638–652, 2025, DOI: 10.59395/ijadis.v6i3.1422.
- [11] F. E. D. Parlian, M. E. Iswanto, and N. S. Wilda, “Analisis Volatilitas dan Segmentasi Rekomendasi Strategis Harga Komoditas Pangan Jawa Barat dengan Pemodelan *Exponential GARCH*,” *Creative Research Journal*, Vol. 11, No. 2, pp. 75–86, 2025.
- [12] C. Liu, L. Zhou, L. Höschle, and X. Yu, “Food Price Dynamics and Regional Clusters: Machine Learning Analysis of Egg Prices in China,” *China Agricultural Economic Review*, Vol. 15, No. 2, pp. 416–432, 2022, DOI: 10.1108/CAER-01-2022-0003.
- [13] L. P. Sari, A. Fanani, and A. H. Asyhar, “Analisis Perbandingan Pengelompokan Kota di Indonesia berdasarkan Indikator Inflasi Tahun 2021 dengan Metode *Ward* dan *K-Means*,” *Jurnal Sains Matematika dan Statistika*, Vol. 9, No. 2, pp. 108–118, 2023, DOI: 10.24014/jsms.v9i2.21100.
- [14] R. L. Manogna, V. Dharmaji, and S. Sarang, “A Novel Hybrid Neural Network-based Volatility Forecasting of Agricultural Commodity Prices: Empirical Evidence from India,” *J. Big Data*, Vol. 12, 2025, DOI: 10.1186/s40537-025-01131-8.
- [15] K. Sheppard, “*Bashtage/arch: Release v7.2.0*,” 2024, Zenodo: v7.2.0. DOI: 10.5281/zenodo.593254.
- [16] Y. Hasan, “Pengukuran *Silhouette Score* dan *Davies-Bouldin Index* pada Hasil Cluster *K-Means* dan *DbSCAN*,” *Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, Vol. 06, No. 01, pp. 60–74, 2024.