

## Clustering Time Series Forecasting Model for Grouping Provinces in Indonesia Based on Granulated Sugar Prices

Fida Fariha Amatullah <sup>1\*</sup>, Erdanisa Aghnia Ilmani <sup>2\*</sup>, Anwar Fitrianto <sup>3\*</sup>, Erfiani <sup>4\*</sup>, L.M. Risman Dwi Jumansyah <sup>5\*</sup>

\* Prodi Statistika dan Sains Data, Sekolah SMI, IPB University, Bogor, 16680, Indonesia

[fidafarihaafida@apps.ipb.ac.id](mailto:fidafarihaafida@apps.ipb.ac.id) <sup>1</sup>, [erdanisaaghnia@apps.ipb.ac.id](mailto:erdanisaaghnia@apps.ipb.ac.id) <sup>2</sup>, [anwarstat@gmail.com](mailto:anwarstat@gmail.com) <sup>3</sup>, [erfiani@apps.ipb.ac.id](mailto:erfiani@apps.ipb.ac.id) <sup>4</sup>, [rismandwijumansyah@apps.ipb.ac.id](mailto:rismandwijumansyah@apps.ipb.ac.id) <sup>5</sup>

### Article Info

#### Article history:

Received 2024-11-10

Revised 2024-12-03

Accepted 2025-01-16

#### Keyword:

*Clustering Time Series, Time Series Analysis, Granulated Sugar Prices, ARIMA Models, Forecasting Models.*

### ABSTRACT

Clustering time series is the process of organizing data into groups based on similarities in specific patterns. This research uses the prices of granulated sugar in each province of Indonesia. According to USDA reports, sugar consumption in Indonesia in 2023 reached 7.9 million tons. On April 26, 2024, the price of granulated sugar peaked in the Papua Mountains at Rp29,320 per kg, while the lowest price was recorded in the Riau Islands at Rp16,460 per kg. The research aims to cluster provinces based on the characteristics of granulated sugar prices and to use forecasting models for each group. Two groups were formed based on the price patterns of granulated sugar over time. The provinces of Papua and West Papua are in group 2, while the other 30 provinces are in group 1. The best model developed using the auto ARIMA method is ARIMA (2, 1, 0), with a MAPE value of 2.36% for cluster 1, and ARIMA (1, 1, 1), with a MAPE value of 2.59% for cluster 2. These values are less than 10%, indicating that the models built using the auto ARIMA method for clusters 1 and 2 are suitable for forecasting.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

### I. PENDAHULUAN

Pengelompokan data deret waktu (*clustering time series*) merupakan proses mengatur data ke dalam kelompok-kelompok berdasarkan kesamaan pola tertentu. Analisis *cluster* bertujuan untuk mengelompokkan objek berdasarkan kemiripan karakteristik. Dalam analisis *cluster*, tidak ada variabel respon, sehingga data dari variabel prediktor digunakan untuk melakukan pengelompokan objek [1]. *Clustering* pada data deret waktu memerlukan pendekatan yang memperhitungkan dimensi temporal, seperti ketergantungan antar titik data dan pola musiman. Sifat data deret waktu yang berkaitan erat dengan urutan waktu sering kali menghasilkan variasi yang tidak mudah diidentifikasi oleh metode *clustering* tradisional seperti *K-Means*, *Hierarchical Clustering* dengan jarak *euclidean*, dan sebagainya yang umumnya digunakan pada data statis [2]. Kajian mengenai pengelompokan data deret waktu telah dilakukan oleh beberapa peneliti. Pengelompokan harga beras di wilayah Indonesia bagian barat dengan pendekatan *clustering time series* dilakukan oleh [3] menghasilkan dua kelompok dengan

menggunakan *hierarchical clustering* beserta ukuran ketidakmiripan *Dynamic Time Warping* (DTW) dan pautan rata-rata, kemudian berdasarkan nilai MAPE, model ARIMA yang dihasilkan pada level *cluster* memiliki keakuratan yang lebih tinggi daripada level provinsi. *Dynamic Time Warping* (DTW) dipilih sebagai ukuran jarak dalam karena kemampuannya yang unggul dalam mengukur kesamaan antar deret waktu dengan memperhitungkan variasi panjang dan pergeseran waktu (*time shift*) pada pola data. Tidak seperti jarak *Euclidean* yang mengukur jarak titik demi titik secara langsung, DTW dapat melakukan penyesuaian (*alignment*) antar titik data yang memiliki fase berbeda, sehingga mampu mengenali pola yang serupa meskipun terjadi pergeseran dalam urutan waktu. Teknik ini memungkinkan identifikasi pola yang tersembunyi atau tidak sejajar dalam waktu, menjadikannya pilihan yang lebih efektif untuk mengelompokkan data deret waktu dengan fluktuasi musiman dan variasi temporal yang kompleks [4]. Kemudian adapun pengelompokan provinsi Indonesia berdasarkan harga beras dengan *hierarchical clustering* menggunakan ukuran ketidakmiripan jarak korelasi Pearson dan pautan rata-rata, hasil

peramalan harga beras level individu dan gerombol cenderung mengikuti data aktual dan memiliki nilai MAPE yang diperoleh tidak melebihi 10%, sehingga pemodelan level kelompok tergolong baik dan representatif untuk dilakukan peramalan [5]. Pengelompokan provinsi di Indonesia berdasarkan harga minyak goreng menggunakan pautan rata-rata dengan ukuran kemiripan *euclidean distance*, menghasilkan koefisien korelasi *cophenetic* sebesar 0,828, model peramalan harga minyak goreng menggunakan pendekatan *Auto ARIMA* menunjukkan akurasi yang baik [6]. Pengelompokan data deret waktu dengan *hierarchical clustering* juga digunakan terhadap kasus positif Covid-19 di DKI Jakarta, yang mencakup 44 distrik. Hasil analisis menunjukkan bahwa pengelompokan menggunakan teknik *Dynamic Time Warping* (DTW) menghasilkan enam kelompok optimal berdasarkan pola distribusi kasus harian [7]. *Clustering time series* yang menggunakan metode *hierarchical clustering* dengan jarak khusus seperti *Dynamic Time Warping* (DTW) terbukti efektif terutama ketika data memiliki variasi temporal yang signifikan [8]. Teknik ini efektif dalam mengelompokkan data deret waktu yang memiliki keterkaitan temporal yang kuat dan pentingnya menggunakan metode jarak seperti DTW dalam mengukur kesamaan pola antar deret waktu dengan variasi kompleks.

Kajian yang telah dilakukan oleh beberapa peneliti sebelumnya menarik peneliti untuk melakukan analisis *clustering time series*. Analisis tersebut akan diterapkan pada data harga gula pasir pada setiap Provinsi di Indonesia. Data harga gula pasir memberikan gambaran yang baik tentang fluktuasi di setiap provinsi, faktor-faktor eksternal seperti kebijakan impor, perubahan nilai tukar rupiah, dan kondisi cuaca ekstrem tidak dimasukkan dalam analisis. Meskipun faktor-faktor ini memiliki potensi pengaruh yang besar, tidak dicantumkan dalam model karena keterbatasan ketersediaan data yang konsisten di seluruh wilayah.

Data harga gula pasir dianalisis dengan memperhatikan tren harga, yaitu perubahan harga secara keseluruhan dari waktu ke waktu, serta fluktuasi musiman, di mana pola harga menunjukkan kenaikan atau penurunan pada periode tertentu seperti bulan puasa atau akhir tahun. Dasar pengelompokan ini bertujuan untuk menangkap variasi pola harga yang tidak dapat diidentifikasi hanya dengan melihat rata-rata harga, tetapi juga mempertimbangkan dinamika temporal dan musiman yang mempengaruhi fluktuasi harga. Pola harga dianalisis menggunakan *Hierarchical Clustering* dengan *Dynamic Time Warping* (DTW) untuk mengakomodasi variasi temporal dan pergeseran pola harga antar provinsi.

Gula pasir adalah salah satu komoditas pangan penting di Indonesia yang dibutuhkan oleh rumah tangga serta industri makanan dan minuman. Menurut laporan USDA, konsumsi gula di Indonesia pada tahun 2023 mencapai 7,9 juta ton. Namun, dengan produksi nasional hanya sekitar 2,6 juta ton, Indonesia harus mengimpor sekitar 5,3 juta ton gula untuk menutupi kekurangan pasokan dalam negeri, baik untuk konsumsi rumah tangga maupun industri [9]. Kondisi ini turut memperumit pasar gula di Indonesia, dengan fluktuasi harga

yang sering terjadi akibat faktor domestik dan internasional, berdampak langsung pada daya beli masyarakat dan stabilitas ekonomi. Fluktuasi harga ini semakin terlihat ketika, menurut laporan CNBC Indonesia, harga gula pasir pada 26 April 2024 mencapai titik tertinggi di Papua Pegunungan, yaitu Rp29.320 per kg, sementara harga terendah tercatat di Kepulauan Riau dengan Rp16.460 per kg. Perbandingan dengan tahun sebelumnya menunjukkan kenaikan harga yang signifikan, di mana pada April 2023, rata-rata harga gula pasir masih berada di kisaran Rp14.400 per kg. Tren kenaikan harga mulai terlihat sejak September 2023, dan mencapai puncaknya pada Desember 2023 dengan harga rata-rata Rp17.270 per kg [10].

Fluktuasi harga gula pasir dapat memberikan dampak signifikan bagi rumah tangga berpenghasilan rendah serta industri makanan dan minuman. Penyebab utamanya sering kali berasal dari faktor eksternal seperti kondisi cuaca yang tidak stabil, serangan hama yang menyebabkan gagal panen, serta infrastruktur yang belum merata di berbagai wilayah. Keterbatasan dalam distribusi membuat harga gula pasir di beberapa daerah di Indonesia lebih tinggi, yang kemudian berkontribusi terhadap inflasi. Ketidakseimbangan antara pasokan yang terbatas dan permintaan yang tinggi semakin memperburuk lonjakan harga komoditas tersebut [11]. Kondisi ini menunjukkan bahwa harga gula pasir di Indonesia tidak hanya dipengaruhi oleh faktor domestik seperti produksi dan distribusi, tetapi juga oleh dinamika pasar internasional. Pemerintah pusat dan daerah memiliki peran penting dalam mengendalikan harga melalui kebijakan yang tepat, terutama dengan memperhatikan perbedaan karakteristik tiap wilayah. Adanya pengelompokan provinsi berdasarkan karakteristik harga gula pasir serta menggunakan model peramalan untuk masing-masing kelompok, kebijakan dapat diimplementasikan dengan lebih efisien, sehingga tidak perlu ada kebijakan yang berbeda untuk setiap provinsi.

Berdasarkan pembahasan sebelumnya, analisis *clustering time series* dilakukan untuk mengelompokkan provinsi di Indonesia berdasarkan pola harga gula pasir bulanan pada tahun 2009 - 2023. Pendekatan ini bertujuan untuk menghasilkan pengelompokan provinsi yang optimal, diikuti dengan peramalan kelompok menggunakan model *auto - ARIMA*. Penggunaan *clustering time series* memberikan efisiensi dalam pemodelan, terutama pada tingkat kelompok provinsi, sehingga lebih praktis dibandingkan melakukan peramalan untuk setiap provinsi secara individu. Selain itu, kombinasi inovatif antara *clustering time series* dan *Auto-ARIMA* ini menawarkan cara baru dalam menganalisis harga gula pasir, dengan hasil yang diharapkan dapat membantu pemerintah dan pemangku kepentingan dalam merancang kebijakan yang lebih adaptif dan berbasis data. Dengan pengelompokan yang lebih akurat, upaya stabilisasi harga dan distribusi gula pasir dapat dilakukan dengan lebih efektif. Pendekatan ini memberikan perspektif baru dalam memahami fluktuasi harga gula pasir, yang sebelumnya lebih sering diterapkan pada komoditas pangan lain seperti beras dan minyak goreng, sehingga memperkaya strategi pengelolaan komoditas strategis di Indonesia.

## II. METODE

### A. Data

Data yang digunakan pada penelitian ini merupakan data bulanan harga gula pasir selama 15 tahun mulai dari 2009-2023 sehingga banyak data yaitu 180 data pada 32 Provinsi di Indonesia. Data diperoleh dari publikasi Badan Pusat Statistik Republik Indonesia (BPS RI) pada periode tertentu, dengan rincian dapat dilihat pada tabel 1.

TABEL I  
DATA DAN SUMBER DATA PENELITIAN

No	Data	Sumber Data
1	Data Bulanan Harga Gula Pasir Tahun 2009	[12]
2	Data Bulanan Harga Gula Pasir Tahun 2010	[13]
3	Data Bulanan Harga Gula Pasir Tahun 2011	[14]
4	Data Bulanan Harga Gula Pasir Tahun 2012	[15]
5	Data Bulanan Harga Gula Pasir Tahun 2013	[16]
6	Data Bulanan Harga Gula Pasir Tahun 2014	[17]
7	Data Bulanan Harga Gula Pasir Tahun 2015	[18]
8	Data Bulanan Harga Gula Pasir Tahun 2016	[19]
9	Data Bulanan Harga Gula Pasir Tahun 2017	[20]
10	Data Bulanan Harga Gula Pasir Tahun 2018	[21]
11	Data Bulanan Harga Gula Pasir Tahun 2019	[22]
12	Data Bulanan Harga Gula Pasir Tahun 2020	[23]
13	Data Bulanan Harga Gula Pasir Tahun 2021	[24]
14	Data Bulanan Harga Gula Pasir Tahun 2022	[25]
15	Data Bulanan Harga Gula Pasir Tahun 2023	[26]

### B. Teknik Analisis

Analisis dilakukan menggunakan metode *clustering time series* dengan *hierarchical clustering*, yang melibatkan beberapa tahapan analisis sebagai berikut:

- Melakukan eksplorasi data untuk mengetahui deskripsi data serta melihat apakah ada pola tren dan musiman pada data.
- Melakukan pengukuran ketidakmiripan untuk pengelompokan berhirarki dengan menggunakan metode *Dynamic Time Warping (DTW)*. DTW adalah algoritma yang bertujuan untuk membandingkan dua deret data waktu dan menghitung lintasan optimal yang meminimalkan jarak antara keduanya [27]. Berbeda dengan algoritma tradisional yang membandingkan urutan diskret atau nilai kontinu secara langsung, DTW

memperluas pendekatan ini. Secara matematis, rumus DTW dirumuskan sebagai berikut:

$$DTW(S, T) = \min_W \left[ \sum_{k=1}^p \delta(w_k) \right] \quad (1)$$

dengan:  $S = s_1, s_2, \dots, s_i, \dots, s_n$  dan  $T = t_1, t_2, \dots, t_j, \dots, t_m$  adalah suatu deret waktu,  $W = w_1, w_2, \dots, w_k$  sebagai jalur lengkungan (*warping path*) yang membentuk matriks ukuran  $n \times m$  [28]. *Warping path* ini membantu menyelaraskan kedua deret waktu tersebut sedemikian rupa sehingga jaraknya minimal. Jalur lengkungan  $W$  terdiri dari pasangan titik  $(i, j)$  yang menghubungkan elemen-elemen dari  $S$  dan  $T$ .

- Melakukan pengelompokan berhirarki menggunakan pautan. Beberapa metode pautan yang diterapkan meliputi [29]:

*Single Linkage*: Menggabungkan kluster berdasarkan jarak terdekat antar elemen dari dua kluster.

$$d_{(xy)z} = \min\{d_{xz}, d_{yz}\} \quad (2)$$

Kuantitas dalam  $d_{xz}$  dan  $d_{yz}$  masing-masing adalah jarak terpendek antara kluster  $x$  dan  $z$  serta antara kluster  $y$  dan  $z$ .

*Complete Linkage*: Menggabungkan kluster berdasarkan jarak terjauh antar elemen dari dua kluster.

$$d_{(xy)z} = \max\{d_{xz}, d_{yz}\} \quad (3)$$

Kuantitas dalam  $d_{xz}$  dan  $d_{yz}$  masing-masing adalah jarak terjauh antara kluster  $x$  dan  $z$  serta antara kluster  $y$  dan  $z$ .

*Average Linkage*: Menghitung rata-rata jarak antar elemen dari dua kluster.

$$d_{(xy)z} = \text{average}\{d_{xz}, d_{yz}\} = \frac{d_{xz} + d_{yz}}{n_{(xy)}n_z} \quad (4)$$

di mana  $n_{(xy)}$  adalah jumlah anggota yang tergabung dalam kluster  $(xy)$  dan  $n_z$  adalah jumlah anggota yang tergabung dalam kluster  $z$  [6].

*Ward's Method*: Menggabungkan kluster dengan meminimalkan varians dalam kluster yang dihasilkan.

$$d_{(xy)z} = \frac{(n_x + n_z)d_{xz} + (n_y + n_z)d_{yz} - n_z d_{xy}}{n_x + n_y + n_z} \quad (5)$$

$n_x$ : jumlah objek dalam kluster  $x$

$n_y$ : jumlah objek dalam kluster  $y$

$n_z$ : jumlah objek dalam kluster  $z$

$d_{xz}$ : jarak antara kluster  $x$  dan kluster  $z$

$d_{yz}$  : jarak antara kluster  $y$  dan kluster  $z$   
 $d_{xy}$  : jarak antara kluster  $x$  dan kluster  $y$

*Centroid Linkage*: Menggabungkan kluster berdasarkan jarak antara centroid (titik tengah) dari masing-masing kluster.

$$\bar{X} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} \tag{6}$$

dimana  $N_1 = N_2$  adalah jumlah objek.

- Evaluasi ukuran ketidakmiripan dilakukan dengan menghitung nilai koefisien korelasi *cophenetic*. Semakin tinggi nilai koefisien korelasi *cophenetic*, semakin baik kualitas ukuran ketidakmiripan yang diterapkan [1]. Dengan persamaan sebagai berikut [30]:

$$c = \frac{\sum_{i < j} (d(i, j) - \bar{d})(v(i, j) - \bar{v})}{\sqrt{[\sum_{i < j} (d(i, j) - \bar{d})^2][\sum_{i < j} (v(i, j) - \bar{v})^2]}} \tag{7}$$

Dengan,  $c$  = koefisien korelasi *Cophenetic*,  $d(i, j)$  = jarak ketidakmiripan antara amatan ke- $i$  dan amatan ke- $j$ ,  $\bar{d}$  = rata-rata jarak ketidakmiripan,  $v(i, j)$  = jarak antara titik dendogram ke- $i$  dan titik dendogram ke- $j$ ,  $\bar{v}$  = rata-rata jarak pada dendogram.

- Menentukan jumlah kluster atau nilai  $k$  yang optimal dilakukan dengan menghitung koefisien *silhouette* dan indeks *calinski harabsz*. Koefisien Silhouette digunakan untuk mengukur seberapa baik objek-objek terkelompok dalam klusternya. Indeks *calinski-harabsz* mengevaluasi rasio antara kepadatan dan pemisahan antar kluster. Nilai  $k$  yang optimal adalah yang menghasilkan nilai koefisien *silhouette* tertinggi dan nilai indeks *calinski-harabsz* terbesar [31].
- Menghitung representasi kluster menggunakan metode rata-rata untuk setiap titik deret waktu dalam satu kluster, diperoleh nilai harga gula pasir yang menjadi representasi kluster tersebut (prototipe).
- Membagi data prototipe menjadi data latih dan data uji.
- Melakukan prediksi prototipe harga gula pasir di setiap kluster menggunakan model ARIMA.
- Mengevaluasi hasil prediksi dan peramalan pada setiap prototipe dengan menggunakan metode MAPE

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{x_t - f_t}{x_t} \right|}{n} \times 100\% \tag{9}$$

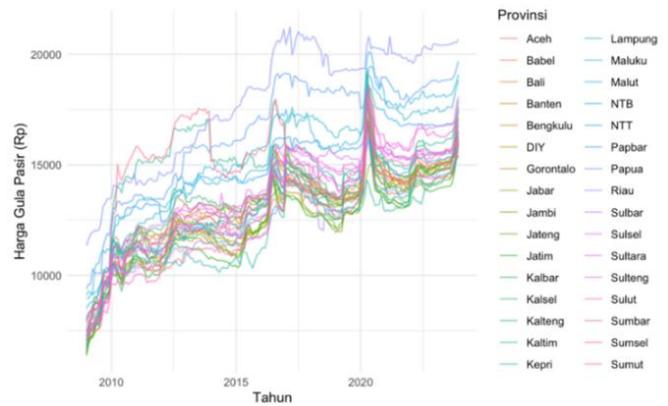
MAPE memiliki range nilai yang dapat dijadikan patokan dalam pengukuran model sebagai berikut [32]:

- < 10% : Kompetensi Model Sangat Baik
- 10 – 20% : Kompetensi Model Baik
- 20 – 50% : Kompetensi Model Layak
- >50% : Kompetensi Model Buruk

### III. HASIL DAN PEMBAHASAN

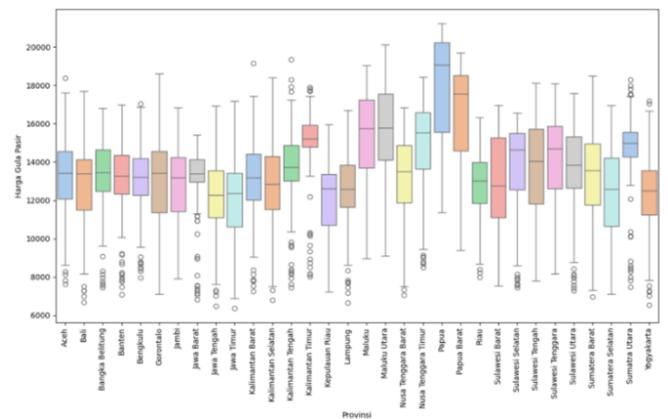
#### A. Eksplorasi Data

Harga gula pasir meningkat di semua Provinsi di Indonesia. Fluktuasi harga gula pasir di setiap Provinsi di Indonesia dapat dilihat pada gambar berikut.



Gambar 1. Perbandingan Harga Gula Pasir Antar Provinsi di Indonesia: 2009-2023

Beberapa provinsi menunjukkan fluktuasi harga yang lebih tajam dibanding provinsi lain. Secara umum, harga gula pasir meningkat di hampir semua provinsi dari tahun 2010 hingga puncak di sekitar 2020. Ada beberapa fluktuasi besar yang terlihat pada pertengahan tahun 2020, kemungkinan disebabkan oleh perubahan pasar atau peristiwa tertentu (misalnya, pandemi), sementara lainnya terlihat lebih stabil. Ini adalah sinyal penting untuk mengelompokkan provinsi berdasarkan pola pergerakan harga.

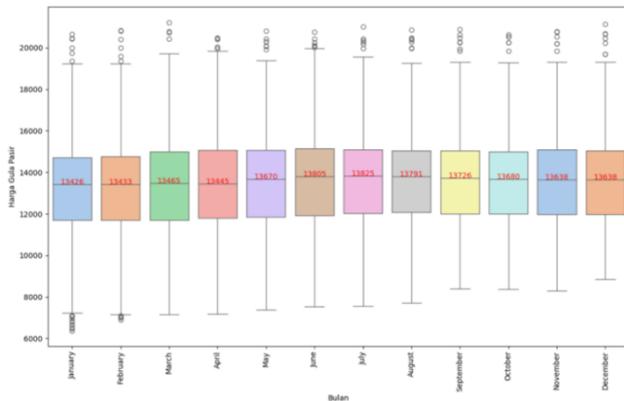


Gambar 2. Boxplot Harga Gula per Provinsi di Indonesia: 2009-2023

Garis horizontal di tengah kotak mewakili median harga gula di setiap provinsi. Sebagian besar provinsi memiliki median harga gula yang berkisar antara 12.000 hingga 16.000 rupiah per kilogram. Provinsi seperti Papua memiliki rentang yang lebih luas, menandakan bahwa harga gula di provinsi tersebut lebih bervariasi dibandingkan dengan provinsi lain seperti Bengkulu dan Banten yang memiliki rentang yang lebih sempit. Provinsi dengan harga gula yang lebih tinggi secara konsisten, seperti Papua, Maluku, dan Kalimantan

Timur. Sementara itu, provinsi seperti Bali, DIY, dan Bengkulu cenderung memiliki harga gula yang lebih rendah.

Titik-titik merah menunjukkan outlier, yaitu harga gula yang sangat berbeda dari harga rata-rata di provinsi tersebut. Beberapa provinsi, seperti Gorontalo, Kalimantan Tengah, dan Sumatera Utara, memiliki banyak outlier, yang bisa menunjukkan adanya fluktuasi harga yang signifikan atau kondisi-kondisi khusus yang mempengaruhi harga gula.

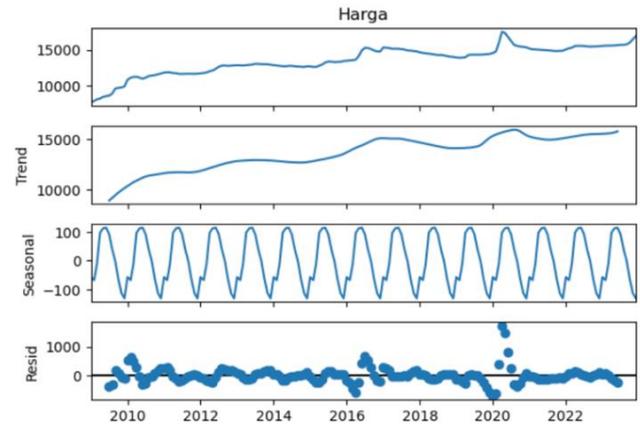


Gambar 3. Pola Musiman Harga Gula di Semua Provinsi di Indonesia: 2009-2023

*Boxplot* pada gambar 3 menunjukkan bahwa rata-rata harga gula di Indonesia berkisar antara Rp13.000. Pada bulan Januari dan Februari, rentang harga relatif lebih rendah, dengan beberapa *outlier* yang sangat rendah. Ini mungkin menunjukkan harga gula yang lebih murah pada awal tahun. Pada bulan Maret hingga Mei, harga gula terlihat sedikit lebih tinggi dibandingkan Januari dan Februari, namun tidak ada perubahan yang signifikan pada nilai tengah. Pola harga cenderung stabil dengan median yang konsisten dan interkuartil yang relatif sama di bulan Juni hingga Desember. Namun, terdapat beberapa outlier yang cukup tinggi, terutama di bulan November dan Desember.

Pencilan yang terlihat pada beberapa bulan (seperti Januari dan Desember) menunjukkan adanya fluktuasi harga yang signifikan pada bulan-bulan tersebut. Fluktuasi harga yang lebih rendah di bulan-bulan tertentu mungkin terkait dengan faktor musiman seperti panen atau produksi gula di dalam negeri atau perubahan kebijakan impor. Kenaikan harga di bulan-bulan tertentu bisa disebabkan oleh permintaan yang meningkat, mungkin terkait dengan hari raya atau musim tertentu di Indonesia.

Harga gula secara umum cenderung meningkat dari tahun 2009 hingga 2023 dengan fluktuasi musiman yang jelas. Terlihat bahwa harga gula mengalami peningkatan bertahap dari tahun 2009 hingga sekitar tahun 2016. Setelah tahun 2016, harga gula relatif stabil dengan beberapa peningkatan di pertengahan tahun 2020 dan sedikit penurunan setelahnya. Ini menunjukkan bahwa ada peningkatan umum dalam harga gula yang mungkin dipengaruhi oleh faktor ekonomi, kebijakan, atau kondisi pasokan dalam jangka panjang.



Gambar 4. Dekomposisi Rata-rata Harga Gula di Indonesia: 2009-2023

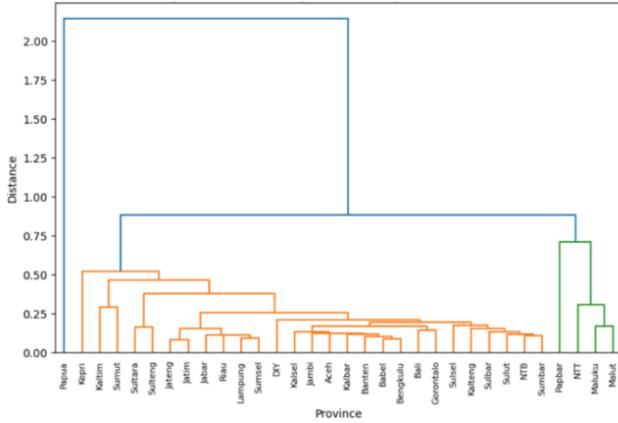
Pola ini menunjukkan bahwa harga gula memiliki fluktuasi musiman yang konsisten setiap tahun. Harga gula cenderung lebih rendah di awal tahun (sekitar bulan Januari hingga Maret) dan lebih tinggi pada pertengahan hingga akhir tahun (sekitar bulan Juli hingga November). Pola musiman ini bisa disebabkan oleh faktor-faktor seperti musim panen, permintaan pasar yang terkait dengan hari raya, atau kebijakan impor gula.

Sisa (*remainder*) mencerminkan fluktuasi harga yang tidak teratur, kejadian acak, atau *outlier* dalam data. Beberapa lonjakan sisa (seperti yang terlihat pada tahun 2014, 2016, dan 2020) menunjukkan adanya kejadian-kejadian luar biasa yang mempengaruhi harga gula secara signifikan, seperti mungkin perubahan kebijakan, bencana alam, atau fluktuasi tajam dalam pasokan atau permintaan. Lonjakan terbesar di tahun 2020 bisa terkait dengan dampak pandemi COVID-19 yang memengaruhi banyak aspek ekonomi, termasuk harga komoditas.

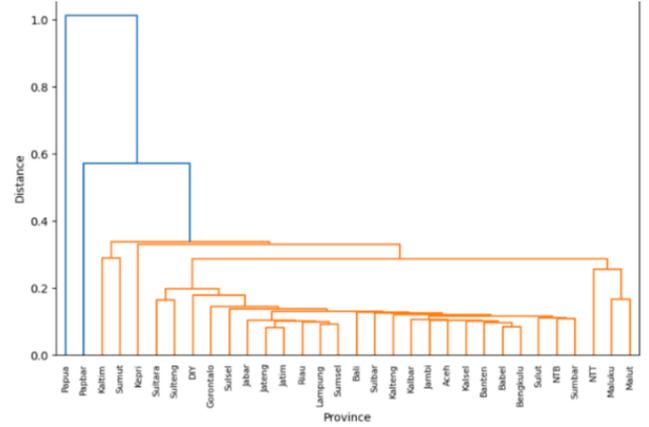
#### B. Clustering Time Series (Pengelompokan Deret Waktu)

Analisis *clustering time series* dilakukan menggunakan pengelompokan berhirarki atau *hierarchical clustering* dengan menghitung jarak antar seri waktu menggunakan *Dynamic Time Warping* (DTW), yang memfasilitasi perbandingan fleksibel antar data meskipun terdapat pergeseran pola waktu. Hasil perhitungan jarak DTW ini digunakan dalam proses *clustering* hierarki, dengan penerapan metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid*, dan *Ward*. Setiap metode menawarkan pendekatan yang berbeda dalam membentuk struktur pengelompokan berdasarkan jarak antar data.

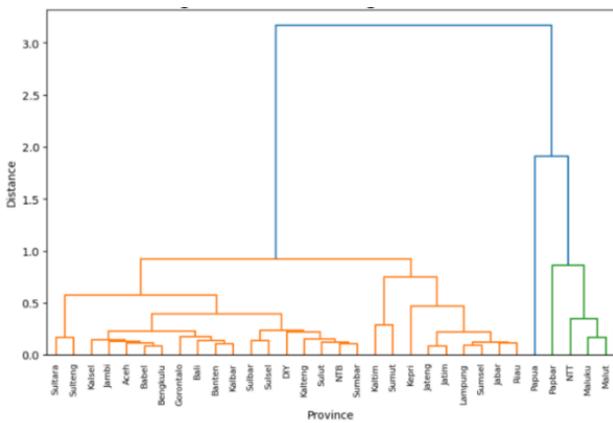
Berdasarkan hasil analisis, grafik dendrogram yang dihasilkan dari proses *clustering* disajikan berikut ini.



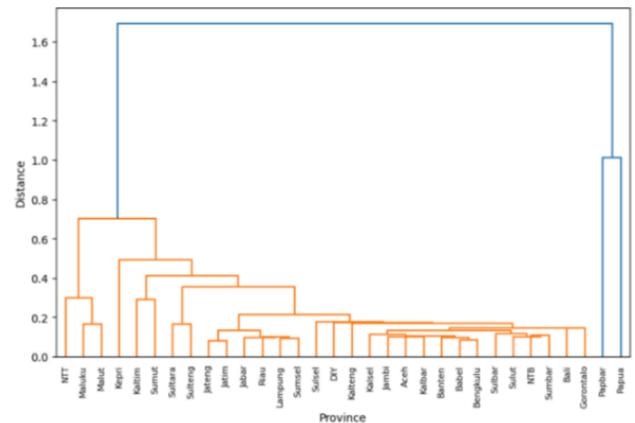
Gambar 5. Dendrogram clustering berhierarki dengan metode Average Linkage.



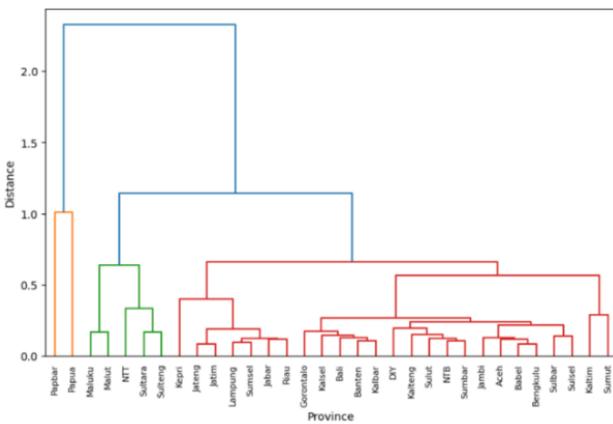
Gambar 8. Dendrogram clustering berhierarki dengan metode Single Linkage



Gambar 6. Dendrogram clustering berhierarki dengan metode Ward



Gambar 9. Dendrogram clustering berhierarki dengan metode Centroid



Gambar 7. Dendrogram clustering berhierarki dengan metode Complete Linkage

Berdasarkan Gambar 8 dan 9 yaitu metode keterpautan “single linkage” dan “centroid” bahwa terdapat 2 kelompok yang terbentuk dari hasil pemotongan dendrogram dengan menggunakan jarak DTW. Selanjutnya dari Gambar 5, 6, dan 7 yaitu metode keterpautan “average linkage”, “ward”, dan “complete linkage” menghasilkan 3 kelompok dengan ukuran jarak DTW.

Selanjutnya, menghitung nilai korelasi cophenetic untuk mengevaluasi metode keterpautan yang terbaik untuk dipilih, hasil perhitungan korelasi cophenetic disajikan pada tabel 2.

TABEL II  
NILAI KORELASI COPHENETIC UKURAN JARAK DTW

	Complete	Single	Average
Cophenetic Correlation	0,76	0,78	0,74
Cophenetic Correlation	0,74	0,77	

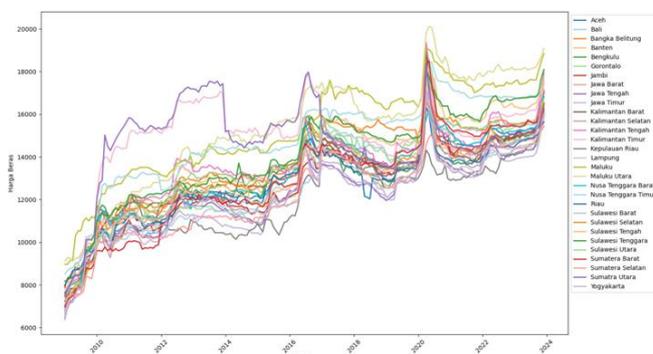
Berdasarkan Tabel 2, nilai korelasi cophenetic tertinggi terdapat pada metode keterpautan “single linkage” yaitu 0.78. Jika korelasi cophenetic mendekati 1, ini menunjukkan bahwa hasil pengelompokkan dengan metode linkage tersebut merepresentasikan jarak asli antar data dengan baik. Artinya, struktur klaster yang dihasilkan mendekati hubungan asli antar data. Sehingga berdasarkan hasil analisis clustering time series dengan metode berhierarki, kelompok yang terbentuk adalah dua kelompok. Kemudian lakukan validasi terhadap jumlah kelompok yang sudah terbentuk dengan menggunakan nilai koefisien silhouette dan indeks calinski harabsz. Hasil validasi disajikan pada Tabel 3.

TABEL III.  
HASIL VALIDASI HIERARCHICAL CLUSTER

Method	Jumlah Klaster	Nilai Silhouette	Calinski Harabasz Index
Ward	2	0,65	41,89
	3	0,61	56,09
	4	0,31	50,35
Average	2	0,75	31,05
	3	0,61	56,09
	4	0,55	48,17
Complete	2	0,74	49,69
	3	0,49	50,43
	4	0,47	47,62
Single	2	0,75	31,05
	3	0,63	31,24
	4	0,28	23,25
Centroid	2	0,74	49,69
	3	0,63	31,24
	4	0,55	48,17

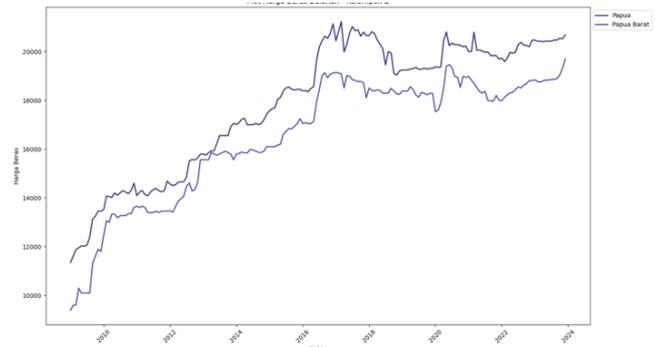
Berdasarkan Tabel 2, hasil analisis menunjukkan bahwa korelasi *cophenetic* tertinggi diperoleh pada metode *Single Linkage* dengan nilai 0.78, yang berarti bahwa metode ini mampu merepresentasikan struktur data dengan baik. Meski demikian, ketika hasil *clustering* divalidasi menggunakan nilai Silhouette dan indeks Calinski-Harabasz, pada Tabel 3, metode keterpautan "centroid" menunjukkan performa yang lebih konsisten. Nilai koefisien Silhouette yang tinggi mengindikasikan bahwa klaster yang terbentuk memiliki kesamaan internal yang baik, sementara indeks Calinski-Harabasz menunjukkan bahwa jarak antar klaster cukup signifikan.

Berdasarkan kombinasi hasil ini, dapat disimpulkan bahwa pendekatan terbaik untuk *clustering time series* menggunakan jarak DTW adalah metode *centroid* dengan dua kelompok. Penggunaan metode ini tidak hanya memberikan pengelompokan yang lebih stabil, tetapi juga menghasilkan struktur klaster yang lebih jelas dan representatif terhadap pola harga antar provinsi. Keputusan untuk memilih dua kelompok atau dua klaster didasarkan pada konsistensi hasil validasi dan kekuatan korelasi *cophenetic* yang menunjukkan representasi yang baik dari hubungan asli antar data. Daerah-daerah pada kedua kelompok disajikan dalam Gambar 10 berikut.



Gambar 10. Plot Harga Gula Pasir - Klaster 1

Pembagian klaster didasarkan pada kesamaan pola temporal harga gula pasir antar provinsi. klaster 1 terdiri dari 30 provinsi yang menunjukkan pola harga relatif stabil, dengan fluktuasi yang mengikuti tren musiman yang umum, seperti kenaikan harga menjelang bulan puasa dan akhir tahun. Harga di klaster ini cenderung lebih konsisten dan mengikuti pola musiman yang serupa, dengan variasi harga yang lebih rendah.



Gambar 11. Plot Harga Gula Pasir - Klaster 2

Sebaliknya, klaster 2 mencakup provinsi Papua dan Papua Barat, yang menunjukkan variabilitas harga yang lebih tinggi dan pola fluktuasi yang lebih tidak teratur. Analisis menunjukkan bahwa provinsi-provinsi dalam kelompok ini mengalami lonjakan harga yang signifikan pada periode tertentu, yang mungkin disebabkan oleh keterbatasan infrastruktur distribusi dan ketergantungan yang lebih besar pada pasokan gula impor. Faktor-faktor ini menyebabkan pola harga yang berbeda secara signifikan dibandingkan dengan provinsi-provinsi di kelompok 1.



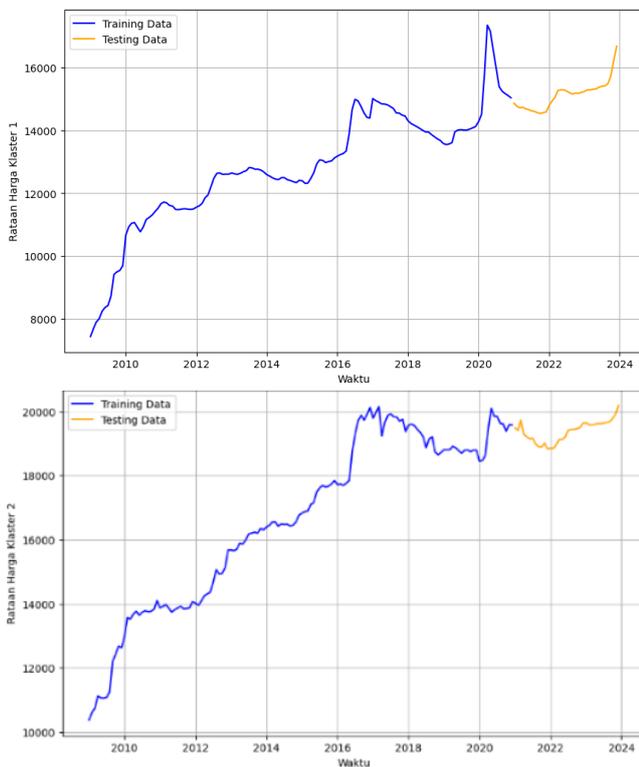
Gambar 12. Distribusi Klaster Harga Gula Pasir di Seluruh Provinsi di Indonesia

Hasil *clustering* harga gula pasir menunjukkan bahwa sebagian besar provinsi di Indonesia masuk dalam satu klaster dengan pola harga yang relatif seragam, kemungkinan karena akses logistik yang baik, ketersediaan pasokan yang stabil, dan permintaan yang moderat. Sementara itu, Papua dan Papua Barat berada di klaster terpisah, yang mungkin disebabkan oleh tantangan distribusi, keterpencilan geografis, dan biaya logistik yang lebih tinggi, sehingga memengaruhi fluktuasi harga gula pasir di wilayah tersebut. Selanjutnya, setiap kelompok akan dibentuk model runtun waktu ARIMA yang diwakili oleh sebuah prototipe, yang bertujuan

merepresentasikan karakteristik anggota dalam kelompok tersebut.

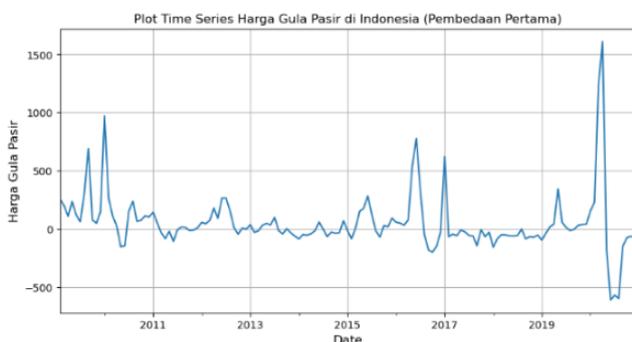
C. Forecasting (Peramalan)

Prototipe diperoleh dengan menghitung rata-rata dari seluruh anggota kelompok, kemudian dilakukan peramalan dengan menggunakan auto ARIMA. Proses peramalan dilakukan dengan membagi data latih dan data uji sebanyak 80:20 yang dapat dilihat pada gambar 11 berikut.



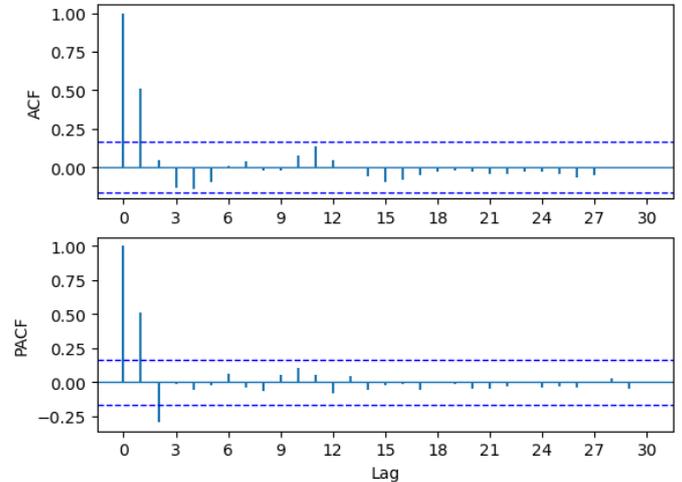
Gambar 13. Grafik Data Latih dan Data Uji untuk Prototipe 1 dan 2

Berdasarkan gambar 13 terlihat bahwa rata-rata prototipe 1 dan prototipe 2 mengalami tren naik, hal ini menunjukkan bahwa data tidak stasioner secara rata-rata. Untuk mengatasi ketidakstasioneran dalam rata-rata, dilakukan differencing data sehingga diperoleh grafik data yang dapat dilihat pada gambar 14 berikut.



Gambar 14. Grafik data prototipe 1 setelah dilakukan differencing 1

Setelah dilakukan differencing terlihat bahwa tidak ada tren pada data, sehingga bisa data telah stasioner, dijelaskan juga oleh nilai Augmented Dickey-Fuller (ADF) sebesar  $4,8 \times 10^{-12}$ . Selanjutnya pembentukan model ARIMA dilakukan dengan mendefinisikan plot ACF dan PACF dapat dilihat sebagai berikut.



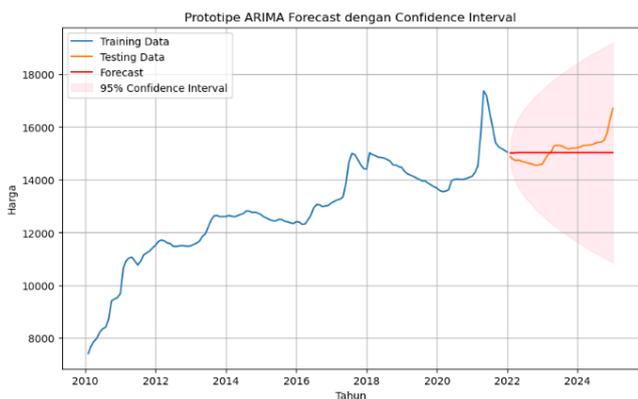
Gambar 15. Plot ACF dan PACF Data Prototipe 1

Model yang dapat terbentuk dari plot ACF dan PACF yaitu ARIMA (0, 1, 1), ARIMA (1, 1, 0), ARIMA (1,1,1), ARIMA (2,1,0), ARIMA (2,1,1). Untuk memilih model terbaik dilakukan pengecekan dengan nilai MAPE, dimana nilai AIC dan BIC terkecil merupakan model terbaik yang dapat digunakan untuk meramalkan data gula pasir untuk prototipe 1. Nilai AIC dan BIC masing-masing model dapat dilihat pada tabel berikut

TABEL IV. NILAI AIC DAN BIC MODEL ARIMA PROTIPE 1

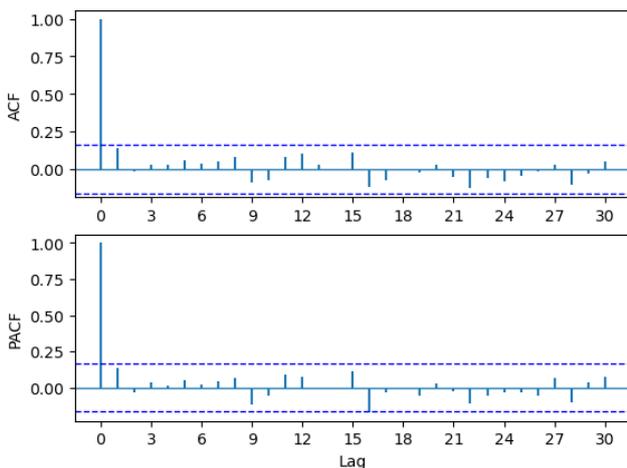
Model	Parameter	P-Value	AIC	BIC
ARIMA (0, 1, 1)	MA1 = 0,57	0,000	1945,08	1951,01
ARIMA (1, 1, 0)	AR1 = 0,53	0,000	1949,87	1955,79
ARIMA (1, 1, 1)	AR1 = 0,28	0,000	1942,44	1951,33
	MA1 = 0,39	0,000		
ARIMA (2, 1, 0)	AR1 = 0,67	0,000	1941,03	1949,92
	AR2 = -0,27	0,000		
ARIMA (2, 1, 1)	AR1 = 0,62	0,104	1943,02	1954,87
	AR2 = -0,24	0,328		
	MA1 = 0,05	0,886		

Model terbaik yang berhasil dibangun dengan proses pengecekan stasioneritas, mendefinisikan berdasarkan plot ACF dan PACF serta melihat nilai AIC dan BIC terkecil yaitu model ARIMA (2, 1, 0). Model tersebut memiliki MAPE sebesar 2,36% yang menyatakan bahwa model baik digunakan untuk peramalan. Hasil ramalan dari model terbaik yang dibangun dapat dilihat pada gambar berikut.



Gambar 16. Grafik Hasil Ramalan Model ARIMA (2,1,0) untuk Prototipe 1

Pembentukan model ARIMA dilanjutkan untuk memodelkan data pada prototipe 2. Grafik data prototipe 2 menunjukkan tren naik yang mengindikasikan bahwa data tidak stasioner dalam rata-rata, sehingga dilakukan *differencing* pertama dan menghasilkan nilai ADF sebesar  $2,98 \times 10^{-18}$  yang menunjukkan bahwa data prototipe 2 telah stasioner setelah dilakukan *differencing* pertama. Setelah data stasioner dapat dilihat plot ACF dan PACF yang terbentuk sebagai berikut.



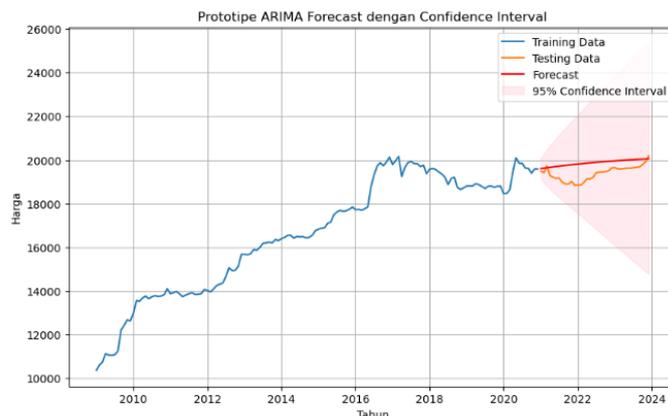
Gambar 17. Plot ACF dan PACF Data Prototipe 2

Berdasarkan plot ACF dan PACF pada gambar 17 dapat membentuk model ARIMA (0, 1, 1), ARIMA (1, 1, 0), dan ARIMA (1, 1, 1). Nilai AIC dan BIC masing-masing model tertera pada tabel berikut.

TABEL V.  
NILAI AIC DAN BIC MODEL ARIMA PROTOTIPE 2

Model	Parameter	P-Value	AIC	BIC
ARIMA (0, 1, 1)	MA1 = 0,20	0,001	1966,58	1972,50
ARIMA (1, 1, 0)	AR1 = 0,20	0,000	1966,28	1972,20
ARIMA (1, 1, 1)	AR1 = 0,97	0,000	1963,93	1972,82
	MA1 = -0,90	0,000		

Model terbaik yang terbentuk berdasarkan data pada prototipe 2 yaitu ARIMA (1, 1, 1). Model ini memiliki nilai MAPE sebesar 2,59%, sehingga model baik digunakan untuk peramalan. Hasil ramalan dari model terbaik sebagai berikut.



Gambar 18. Grafik Hasil Ramalan Model ARIMA (1, 1, 1) Prototipe 2

Bentuk lengkungan berwarna merah muda pada gambar 16 dan 18 merupakan nilai selang kepercayaan atau *confidence interval* yang berguna untuk melihat rentang perkiraan letak nilai sebenarnya. Berikut nilai selang kepercayaan dari prototipe 1 yang disajikan dalam Tabel 6.

TABEL VI.  
HASIL SELANG KEPERCAYAAN PROTOTIPE 1

Waktu	Periode	Prediksi	Batas Bawah	Batas Atas
01/01/21	1	15015,38	14602,38	15428,37
01/02/21	2	15013,10	14208,61	15817,58
01/03/21	3	15020,02	13908,46	16131,58
...	...	...	...	...
...	...	...	...	...
01/10/23	34	15025,78	10987,96	19063,60
01/11/23	35	15025,78	10928,50	19123,06
01/12/23	36	15025,78	10869,89	19181,67

TABEL VII.  
HASIL SELANG KEPERCAYAAN PROTOTIPE 2

Waktu	Periode	Prediksi	Batas Bawah	Batas Atas
01/01/21	1	19603,80	19156,35	20051,25
01/02/21	2	19624,02	18968,66	20279,37
01/03/21	3	19643,66	18813,65	20473,67
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
01/10/23	34	20041,02	15015,93	25066,11
01/11/23	35	20048,85	14896,42	25201,29
01/12/23	36	20056,47	14777,19	25335,74

Model memprediksi harga untuk periode pertama sebesar 15015,38. Namun, mempertimbangkan ketidakpastian dalam prediksi, rentang harga yang mungkin terjadi berkisar antara

14602,38 (batas bawah) hingga 15428,37 (batas atas). Artinya, harga aktual diperkirakan berada dalam rentang ini dengan tingkat kepercayaan 95%, begitupun untuk periode selanjutnya. Hasil selang kepercayaan untuk prototipe 2 disajikan dalam Tabel 7.

Model memprediksi harga untuk periode pertama sebesar 19603,80. Namun, dengan mempertimbangkan ketidakpastian dalam prediksi, rentang harga yang mungkin terjadi berkisar antara 19156,35 (batas bawah) hingga 20051,25 (batas atas). Artinya, harga aktual diperkirakan berada dalam rentang ini dengan tingkat kepercayaan 95%, begitupula untuk periode selanjutnya.

#### IV. KESIMPULAN

Penelitian ini memberikan hasil bahwa metode clustering time series mampu membentuk kelompok berdasarkan pola deret waktu untuk data harga gula pasir pada setiap Provinsi di Indonesia. Sebanyak 2 kelompok berhasil dibangun menggunakan metode centroid linkage yang memiliki nilai sillhouette sebesar 0,74 dan nilai Calinski Harabasz sebesar 49,69. Provinsi Papua dan Papua Barat berada dalam kelompok 2 sedangkan 30 Provinsi lainnya ada dalam kelompok 1. Model terbaik yang berhasil dibangun menggunakan metode auto ARIMA yaitu ARIMA (2, 1, 0) dengan nilai MAPE sebesar 2,36% untuk *cluster* 1 dan ARIMA (1, 1, 1) dengan nilai MAPE sebesar 2,59% untuk *cluster* 2. Nilai tersebut kurang dari 10% yang menyatakan bahwa model yang dibangun menggunakan metode ARIMA untuk *cluster* 1 dan 2 baik digunakan untuk peramalan.

#### DAFTAR PUSTAKA

- [1] D. A. N. Sirodj, I. M. Sumertajaya, and A. Kurnia, "Analisis Clustering Time Series untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Manusia Jenis Kelamin Perempuan," *Stat. J. Theor. Stat. Its Appl.*, vol. 23, no. 1, pp. 29–37, 2023, doi: 10.29313/statistika.v23i1.2181.
- [2] L. Li and B. A. Prakash, "Time series clustering: Complex is simpler!," *Proc. 28th Int. Conf. Mach. Learn. ICML 2011*, pp. 1–8, 2011.
- [3] S. U. Wijaya and N. N. Ngatini, "Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series," *Limits J. Math. Its Appl.*, vol. 17, no. 1, p. 51, 2020, doi: 10.12962/limits.v17i1.5994.
- [4] P. Esling and C. Agon, "Time-series data mining," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 12–34, 2012, doi: 10.1145/2379776.2379788.
- [5] M. Ulinuha, F. M. Afendi, and I. M. Sumertajaya, "Study of Clustering Time Series Forecasting Model for Provincial Grouping in Indonesia Based on Rice Price," *Indones. J. Stat. Its Appl.*, vol. 6, no. 1, pp. 50–62, 2022, doi: 10.29244/ijsa.v6i1p50-62.
- [6] M. A. Zen, S. Wahyuningsih, and A. T. R. Dani, "Aplikasi Pendekatan Agglomerative Hierarchical Time Series Clustering untuk Peramalan Data Harga Minyak Goreng di Indonesia," *Semin. Nas. Off. Stat.*, pp. 293–302, 2022, doi: 10.34123/semnasoffstat.v2022i1.1394.
- [7] M. Yohansa, K. A. Notodiputro, and E. Erfiani, "Dynamic Time Warping Techniques for Time Series Clustering of Covid-19 Cases in DKI Jakarta," *ComTech Comput. Math. Eng. Appl.*, vol. 13, no. 2, pp. 63–73, 2022, doi: 10.21512/comtech.v13i2.7413.
- [8] J. Paparrizos and L. Gravano, "K-Shape: Efficient and Accurate Clustering of Time Series," *SIGMOD Rec.*, vol. 45, no. 1, pp. 69–76, 2016, doi: 10.1145/2949741.2949758.
- [9] Sahara, F. A. D. Putro, L. A. Putri, and G. Prawira, "Dinamika Pasar Gula Global dan Lokal, Manis atau Pahit?," *fem.ipb.ac.id*, 2024. <https://fem.ipb.ac.id/index.php/2024/02/12/dinamika-pasar-gula-global-dan-lokal-manis-atau-pahit/> (accessed Sep. 12, 2024).
- [10] Damiana, "Harga Gula Makin Tak Terkendali, Hari Ini Pecah Rekor Tembus Rp18.200," *cncbincindonesia.com*, 2024. <https://www.cncbincindonesia.com/news/20240426131817-4-533734/harga-gula-makin-tak-terkendali-hari-ini-pecah-rekor-tembus-rp18200> (accessed Sep. 12, 2024).
- [11] T. Santoso and M. U. Basuki, "Aplikasi Model Garch Pada Data Inflasi Bahan Makanan Indonesia Periode 2005.1- 2010.6," *J. Organ. dan Manaj.*, vol. 7, no. 1, pp. 38–52, 2011, [Online]. Available: [www.bps.go.id](http://www.bps.go.id).
- [12] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2010*. Jakarta: Badan Pusat Statistik, 2010.
- [13] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2011*. Jakarta: Badan Pusat Statistik, 2011.
- [14] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2012*. Jakarta: Badan Pusat Statistik, 2012.
- [15] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2013*. Jakarta: Badan Pusat Statistik, 2013.
- [16] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan (Data 2013)*. Jakarta: Badan Pusat Statistik, 2014.
- [17] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2014*. Jakarta: Badan Pusat Statistik, 2015.
- [18] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2015*. Jakarta: Badan Pusat Statistik, 2016.
- [19] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2016*. Jakarta: Badan Pusat Statistik, 2017.
- [20] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2017*. Jakarta: Badan Pusat Statistik, 2018.
- [21] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2018*. Jakarta: Badan Pusat Statistik, 2019.
- [22] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2019*. Jakarta: Badan Pusat Statistik, 2020.
- [23] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2020*. Jakarta: Badan Pusat Statistik, 2021.
- [24] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2021*. Jakarta: Badan Pusat Statistik, 2022.
- [25] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2022*. Jakarta: Badan Pusat Statistik, 2023.
- [26] Badan Pusat Statistik, *Statistik Harga Konsumen Perdesaan Kelompok Makanan 2023*. Jakarta: Badan Pusat Statistik, 2024.
- [27] A. D. Munthe, "Penerapan Clustering Time Series Untuk Menggerombolkan Provinsi Di Indonesia Berdasarkan Nilai Produksi Padi," *J. Litbang Sukowati Media Penelit. dan Pengemb.*, vol. 2, no. 2, p. 11, 2019, doi: 10.32630/sukowati.v2i2.61.
- [28] D. Astuti, D. Y. Hartanti, S. T. Nurhayanti, and H. Fransiska, "Clustering and Forecasting of Covid-19 Data in Indonesia," *J. Mat. Stat. dan Komputasi*, vol. 18, no. 3, pp. 324–335, 2022, doi: 10.20956/j.v18i3.18882.
- [29] D. T. Utari and D. S. Hanun, "Hierarchical Clustering Approach for Region Analysis of Contraceptive Users," *EKSAKTA J. Sci. Data Anal.*, vol. 2, no. 2, pp. 99–108, 2021, doi: 10.20885/eksakta.vol2.iss2.art3.
- [30] S. Saraçlı, N. Doğan, and I. Doğan, "Comparison of hierarchical cluster analysis methods by cophenetic correlation," *J. Inequalities Appl.*, no. December, pp. 1–8, 2013, doi: 10.1186/1029-242X-2013-203.
- [31] Scikit Learn, "Clustering," <https://scikit-learn.org/>, 2024. <https://scikit-learn.org/stable/modules/clustering.html> (accessed Oct. 01, 2024).
- [32] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Averagedan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *J. Sist. dan Inform.*, vol. 13, no. 2, pp. 36–45, 2019.