

## Study of Clustering Time Series Forecasting Model for Provincial Grouping in Indonesia Based on Rice Price\*

Kajian Model Peramalan Clustering Time Series untuk Penggerombolan Provinsi Indonesia berdasarkan Harga Beras

Muhammad Ulinnuha<sup>1</sup>, Farit M Afendi<sup>2‡</sup> and I Made Sumertajaya<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, IPB University, Indonesia

<sup>‡</sup>corresponding author: [fmafendi@apps.ipb.ac.id](mailto:fmafendi@apps.ipb.ac.id)

Copyright © 2022 Muhammad Ulinnuha, Farit M Afendi and I Made Sumertajaya. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Most Indonesians consume rice as the main staple. The high low price of rice has an impact on farmers and communities, especially those who cannot afford it. Rice price forecasting is one of the important information to be considered for future rice prices. The data used is secondary data sourced from bps publication, Rural Consumer Price Statistics: Food Group, from January 2008 to December 2019 for 32 provinces in Indonesia. Time series modeling and forecasting is usually done on a single variable using ARIMA. However, modeling becomes inefficient if there are many variables, so clustering time series analysis is performed using correlation distance with the clustering method of average linkage hierarchy. Cluster level ARIMA modeling with 4 clusters provides high efficiency because only by doing 4 times modeling results in accuracy values not much different from individual level modeling. The results obtained by individual-level ARIMA Modeling resulted in an average MAPE of 3.36%, while cluster-level ARIMA modeling with 4 clusters resulted in an average MAPE value of 4.27%, with a second MAPE difference of -0.91%. Formally conducted z test, the results obtained there is no difference between individual-level MAPE and cluster-level MAPE. This means that cluster-level modeling is relatively good and representative.

**Keywords:** ARIMA, clustering time series, correlation distance, rice price.

---

\* Received: Jun 2021; Reviewed: Apr 2022; Published: May 2022

## 1. Pendahuluan

Beras merupakan salah satu komoditas pangan yang mempengaruhi kesejahteraan masyarakat di Indonesia. Mayoritas penduduk Indonesia memenuhi kebutuhan hidupnya dengan mengkonsumsi beras sebagai makanan pokok. Beras merupakan komoditas dengan permintaan inelastis, artinya perubahan harga tidak menyebabkan perubahan jumlah permintaan konsumen dan apabila ketersediaan berkurang akan mengakibatkan harga melambung tinggi sehingga dapat menyebabkan tidak terjangkau oleh konsumen (Siswanto et al., 2018).

Dari beberapa aspek ekonomi pangan, harga merupakan salah satu aspek penting yang perlu mendapatkan perhatian dan perlu adanya kebijakan khusus yang mengatur mengenai harga pangan salah satunya ialah harga beras. Jika harga tinggi dapat mengakibatkan kekhawatiran adanya rawan pangan, sebaliknya jika harga beras rendah akan mengurangi kesejahteraan petani. Peramalan dapat digunakan untuk mencari informasi yang penting diharapkan mampu untuk memberikan harga terbaik yang tidak merugikan dari pihak tertentu agar mencapai swasembada pangan.

Salah satu teknik peramalan yang digunakan ialah pemodelan. Pemodelan data deret waktu dengan provinsi yang banyak dapat dilakukan secara individu [satu persatu]. Akan tetapi, banyaknya provinsi yang digunakan menyebabkan proses pemodelan secara individu tersebut menjadi tidak efisien, agar menjadi lebih efisien dapat dilakukan penataan provinsi terlebih dahulu kedalam suatu gerombol. Penggerombolan dapat dikatakan baik jika presisi nilai prediksi dan ramalan yang diperoleh sama baiknya dengan pemodelan yang dilakukan secara individu.

Beberapa penelitian peramalan harga beras yang pernah dilakukan oleh Mardianto et al. (2020) mengenai peramalan harga beras pada amazon forecast dan sagemaker menggunakan ARIMA. Penelitian terkait penggerombolan pada data deret waktu telah banyak dilakukan. Maharaj dan Inder (1999) membandingkan peramalan pada peubah ganda secara satu persatu dengan peramalan pada saat peubah-peubah tersebut dilakukan penggerombolan terlebih dahulu. Penggerombolan dilakukan berdasarkan kesamaan model dan diperoleh hasil yang tidak berbeda dengan peramalan secara satu persatu. Kumar (2008) juga melakukan penelitian dengan mengkombinasikan penggerombolan dan peramalan. Wijaya dan Ngatini (2020) melakukan pengembangan pemodelan harga beras di wilayah Indonesia bagian barat dengan pendekatan *Clustering Time Series*. Utami (2018) mengkaji model VARX untuk peramalan inflasi menurut sub kelompok komoditi di Jakarta dengan pendekatan *tsclust* sebagai *preprocessing* dengan beberapa jarak yang digunakan salah satunya jarak korelasi.

D'Urso dan Maharaj (2009) yang melakukan penggerombolan deret waktu berdasarkan *Autocorrelation Function* (ACF) pada masing-masing objek untuk mengukur kesamaan dari deret waktu, dengan pertimbangan ACF yang mampu menggambarkan struktur deret waktu dengan baik dan sifatnya dinamis mengikuti waktu. Pendekatan berbasis feature ini ditujukan untuk merepresentasikan struktur dinamis setiap deret dengan fitur dimensi yang lebih rendah, sehingga memungkinkan reduksi data deret waktu pada dasarnya yang berdimensi tinggi dengan cara feature series yang diekstrak menggunakan ACF sehingga penggerombolannya menggunakan nilai ACF pada masing-masing objek agar prosedur penggerombolan yang lebih efisien Montero dan Vilar (2015).

Metode penggerombolan dibagi kedalam dua gerombol hierarki dan tidak hierarki. Hierarki yaitu metode penggerombolan dengan banyaknya gerombol belum diketahui dan tidak hierarki adalah penggerombolan dengan banyaknya gerombol diketahui Rencher (2005). Novidiyanto dan Dani (2020) melakukan analisis gerombol data deret

waktu kasus aktif pada covid-19 berdasarkan provinsi di Indonesia. Munthe (2019); Novidianto dan Dani (2020) menggunakan analisis gerombol hierarki dengan beberapa *linkage*, hasil diperoleh menunjukkan *average linkage* memberikan hasil terbaik sehingga metode *average linkage* lebih stabil dibandingkan *linkage* lainnya.

Berdasarkan penjelasan sebelumnya mengenai analisis gerombol data deret waktu dengan menggunakan ukuran ketidakmiripan yang berbeda serta pemodelan dan peramalan sesuai dengan perilaku data, untuk itu penelitian akan mengkaji penggerombolan provinsi berdasarkan pola harga beras konsumen perdesaan menggunakan metode hierarki dengan ukuran ketidakmiripan yang digunakan jarak korelasi. Salah satu model peramalan yang dapat digunakan ialah ARIMA sehingga untuk melakukan peramalan berdasarkan hasil dari penggerombolan pada penelitian ini menggunakan model ARIMA, selanjutnya untuk mengetahui kinerja pemodelan ARIMA level gerombol baik atau tidak dapat dilakukan pemodelan ARIMA level individu. Karakteristik dari penelitian ini yang belum pernah dilakukan sebelumnya melakukan penggerombolan harga beras berdasarkan provinsi di Indonesia dengan data yang sudah disesuaikan kemudian dilakukn penggerombolan berdasarkan dari karakteristik data atau *autocorrelation function*.

## 2. Metodologi

### 2.1 Bahan dan Data

Data pada penelitian ini menggunakan data sekunder dari publikasi BPS, statistik harga konsumen perdesaan: kelompok makanan, mulai Januari 2008 sampai Desember 2019 untuk 32 provinsi di Indonesia dimana data harga provinsi DKI Jakarta dan Kalimantan Utara untuk periode tersebut belum tersedia. Pengumpulan data harga beras konsumen perdesaan dilakukan secara rutin setiap bulan oleh BPS melalui survei harga konsumen perdesaan, melalui wawancara secara langsung ke pedagang eceran yang berada di pasar perdesaan secara *purposive*. Harga beras yang digunakan pada penelitian ini merupakan rata-rata dari berbagai jenis atau merek beras yang tersedia di wilayah tersebut. Penghitungan harga beras di masing-masing provinsi setiap bulannya menggunakan formula rata-rata geometrik (BPS, 2018).

### 2.2 Metode Penelitian

Prosedur penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Eksplorasi data menggunakan plot, data Januari 2008 sampai Desember 2019 yaitu dengan melihat pergerakan perilaku data harga beras di provinsi sentra dan tidak sentra produksi beras.
2. Membagi data menjadi data latih dan data uji, data latih ini akan digunakan untuk keperluan analisis gerombol data deret waktu dan pemodelan, sedangkan data uji digunakan dalam evaluasi hasil peramalan. Data latih yang digunakan yaitu data harga beras pada periode Januari 2008- Desember 2017 sedangkan data uji periode Januari 2018 - Desember 2019.
3. Memeriksa kestasioneran data harga beras pada masing-masing 32 provinsi dengan mengamati pola *correlogram* ACF dan PACF dan uji *Augmented Dickey-Fuller (ADF)*, dengan persamaan sebagai berikut :

$$\nabla y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^k \phi_j \nabla y_{t-j} + \varepsilon_t \quad [1]$$

dimana  $\alpha$  adalah intersep,  $\beta t$  adalah trend,  $\phi_j$  adalah koefisien autoregressive,  $\gamma$  adalah koefisien regresi dan  $\varepsilon_t$  adalah galat (Kleiber dan Zeileis, 2008).

4. Melakukan transformasi *logaritma natural* ( $\ln$ ) jika terdapat data tidak stasioner dalam ragam dan proses *differencing* jika terdapat data yang tidak stasioner dalam nilai tengah, hingga diperoleh data yang stasioner.
5. Dilakuan pemodelan pada level individu menggunakan *Autoregressive Integrated Moving Average* [ARIMA]. Bentuk umum ARIMA[ $p,d,q$ ] sebagai berikut (Wei, 2006).

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)a_t \tag{2}$$

dengan  $B$  disebut sebagai operator *backshift*,  $\theta$  adalah parameter *moving average*,  $BX_t = X_{t-1}$  dan  $(1-B)^d = \nabla^d$ ;  $\phi$  merupakan parameter *autoregressive*; sementara  $a_t$  adalah nilai galat pada saat- $t$ . Sedangkan jika terdapat musiman pada data, model ARIMA[ $p,d,q$ ][ $P,D,Q$ ]s musiman sebagai berikut :

$$\Phi(B^S)\phi(B)(1-B)^d(1-B^S)^D X_t = \theta(B)\Theta(B^S)a_t \tag{3}$$

dengan  $S$  adalah periode musiman,  $\Theta$  adalah parameter *moving average* musiman,  $\Phi$  adalah parameter *autoregressive* musiman.

6. Identifikasi model dengan melihat plot ACF dan PACF untuk memperoleh ordo  $p$ ,  $d$  dan  $q$  untuk non musiman, jika musiman menentukan ordo  $P,D$  dan  $Q$ . ACF juga dapat digunakan untuk mengidentifikasi orde  $q$  pada metode peramalan sedangkan PACF juga dapat digunakan untuk mengidentifikasi ordo  $p$  pada metode peramalan *Box-Jenkins* (Cryer dan Chan, 2008).

Tabel 1 . Identifikasi ACF dan PACF

Model	ACF	PACF
AR[ $p$ ]	<i>Tails off</i>	<i>Cuts Off setelah lag p</i>
MA[ $q$ ]	<i>Cuts Off setelah lag q</i>	<i>Tails off</i>
ARMA[ $p,q$ ]	<i>Tails off</i>	<i>Tails off</i>

7. Melakukan *overfitting* dengan menambahkan 1 parameter AR dan atau MA secara bergantian pada ordo hasil identifikasi.
8. Pendugaan parameter dengan metode *maximum likelihood estimation*.
9. Penentuan model terbaik dari seluruh kandidat dan *overfitted* model adalah model dengan nilai *AICc* terkecil . *AICc* mengkoreksi jumlah kuadrat galat dengan banyaknya parameter dengan tujuan untuk mengurangi bias (Montgomery et al., 2015).
10. Pengujian diagnostik sisaan terhadap model terbaik yang diperoleh. Model yang baik dan layak ialah model yang memiliki sisaan yang saling bebas dengan uji *Ljung-Box*, sisaan menyebar normal dengan uji *Jarque-Bera* (Montgomery et al., 2015).
11. Melakukan prediksi harga beras di tiap provinsi.
12. Evaluasi hasil prediksi harga beras di tiap provinsi dengan data latih menggunakan MAPE. dapat dihitung berdasarkan bersamaan berikut (Montgomery et al., 2015).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100 \quad [4]$$

$x_t$  = Data aktual pada periode ke-t dan  $\hat{x}_t$  = Data prediksi pada periode ke-t

Penggerombolan dan pemodelan data deret waktu dengan tahapan sebagai berikut:

13. Pada penelitian ini jenis pendekatan penggerombolan data deret waktu ialah berbasis karakteristik data (*eatured-based*). Pendekatan berbasis fitur ini ditujukan untuk merepresentasikan struktur dinamis setiap deret dengan fitur dimensi yang lebih rendah, sehingga memungkinkan reduksi data deret waktu pada dasarnya yang berdimensi tinggi dengan cara fitur serial yang diekstrak menggunakan *autocorrelation function* (ACF) sehingga penggerombolannya menggunakan nilai ACF dimasing-masing gerombol agar prosedur penggerombolan efisien (Montero dan Vilar, 2015).
14. Menghitung jarak atau ukuran ketidakmiripan pada data dengan ukuran ketidakmiripan yang digunakan yaitu jarak korelasi. dengan persamaan sebagai berikut (Golay et al., 1998):

$$COR(x_t, v_t) = \frac{\sum_{t=i}^T (X_t - \bar{X}_T)(V_t - \bar{V}_T)}{\sqrt{\sum_{t=i}^T (X_t - \bar{X}_T)^2} \sqrt{\sum_{t=i}^T (V_t - \bar{V}_T)^2}} \quad [5]$$

dimana  $\bar{X}_T$  dan  $\bar{V}_T$  adalah rata-rata dari  $x_t$  dan  $v_t$ , dengan matriks jarak sebagai berikut :

$$d_{COR.1}(x_t, v_t) = \sqrt{2(1 - COR(x_t, v_t))} \quad [6]$$

15. Melakukan penggerombolan berhirarki dengan metode pautan rata-rata [*average linkage*]. Penelitian ini menggunakan metode pautan rata-rata. Adapun definisi metode pautan rata-rata sebagai berikut (Rencher, 2005).

$$d(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i y_j) \quad [7]$$

dengan  $d[A, B]$  adalah jarak antar objek  $y_i$  dalam gerombol A dan objek  $y_j$  dalam gerombol B sedangkan  $n_A, n_B$  merupakan banyaknya objek pada gerombol A dan B.

16. Mengevaluasi ukuran ketidakmiripan dengan cara menghitung nilai koefisien korelasi *chopenetic*. Nilai koefisien korelasi *cophenetic* semakin besar, maka semakin baik ukuran ketidakmiripan yang digunakan (Sokal dan Rohlf, 1962).
17. Menentukan banyaknya gerombol atau nilai  $k$  optimum dengan cara menghitung koefisien *silhouette* dan rasio keragaman. semakin tinggi nilai *silhouette*

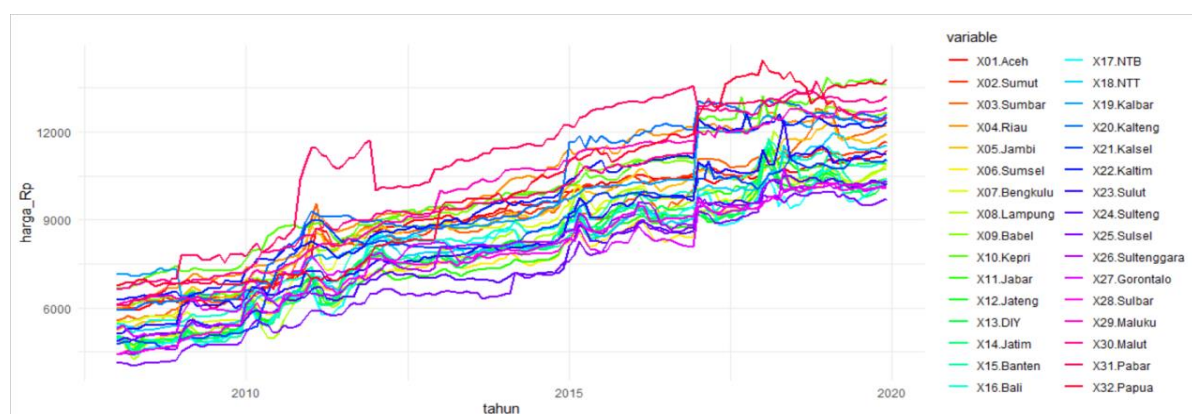
menunjukkan kualitas penggerombolan yang baik (Kaufman dan Rousseeuw, 2009).

18. Selain koefisien *silhouette* pertimbangan dalam penentuan banyak gerombol yang optimal menggunakan perbandingan simpangan baku di dalam gerombol ( $S_w$ ) dengan simpangan baku antar gerombol ( $S_b$ ) sebagai ukuran keragaman dalam gerombol dan keragaman antar gerombol (Kalkstein et al., 1987).
19. Mengidentifikasi anggota gerombol pada masing-masing  $k$  yang terpilih.
20. Setelah menghitung perwakilan gerombol dengan pendekatan teknik menghitung rata-ran untuk setiap titik deret waktu pada satu gerombol, sehingga diperoleh nilai harga beras yang merupakan perwakilan dari gerombol tersebut (*prototype*), yang nantinya akan dilakukan proses pemodelan menggunakan ARIMA pada level gerombol. Pemodelan ARIMA pada level gerombol juga menggunakan prosedur iterasi *Box – Jenkins* seperti ARIMA pada level individu.
21. Melakukan prediksi *prototype* harga beras di setiap gerombol.
22. Evaluasi hasil prediksi dan ramalan *prototype* menggunakan MAPE.
23. Membandingkan hasil prediksi dan ramalan pemodelan dengan dan tanpa penggerombolan menggunakan MAPE dengan melakukan 3 kali ulangan.

### 3. Hasil dan Pembahasan

#### 3.1 Eksplorasi Data

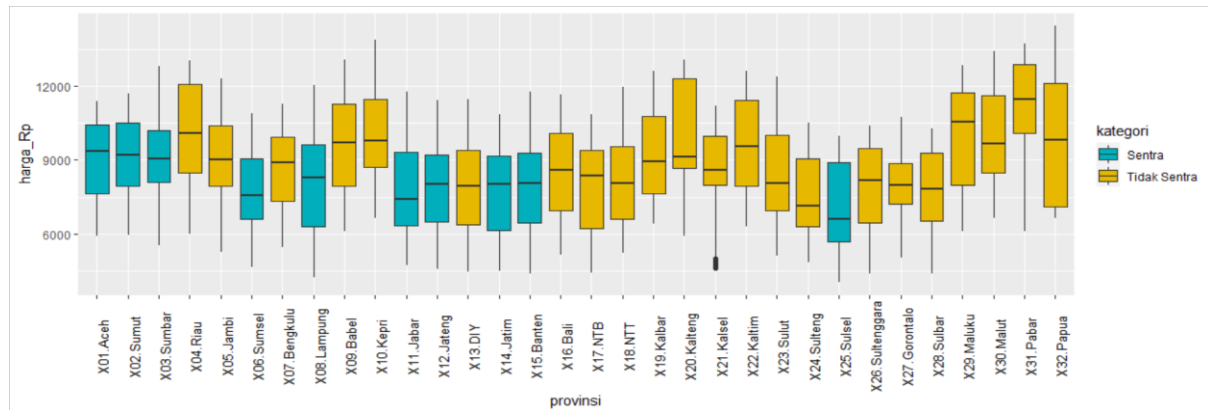
Eksplorasi data dilakukan dengan membuat grafik data deret waktu untuk melihat pergerakan atau pola data harga beras dari waktu ke waktu. Di Indonesia terdapat provinsi yang menghasilkan produksi beras, menurut kementerian pertanian republik Indonesia, dengan metode Kerangka Sampling Area (KSA) yang dilakukan oleh BPS terdapat 10 provinsi di Indonesia sebagai produsen beras tertinggi pada tahun 2019, 10 provinsi tersebut akan di kategorikan sebagai provinsi sentra beras. Selain itu juga akan dilakukan eksplorasi data harga beras berdasarkan harga beras menurut kategori sentra beras dan tidak sentra beras.



Gambar 1 : Plot Harga Beras

Pola harga beras menurut provinsi pada periode tahun 2008-2019 secara umum memiliki pola hampir serupa yaitu setiap tahunnya mengalami kenaikan terdapat beberapa provinsi mengandung pola musiman, beberapa provinsi mengalami pergerakan harga beras yang berbeda dari provinsi lainnya seperti yang terlihat provinsi Papua Barat memiliki harga beras yang relatif tinggi dari pada provinsi lainnya, sekitar tahun 2011 hingga 2012 provinsi Papua Barat mengalami kenaikan harga beras cukup tinggi dibandingkan tahun-tahun sebelumnya. Selain itu harga beras relatif tinggi kebanyakan berada di Indonesia bagian timur sementara harga beras relatif

rendah cenderung berada di pulau jawa, karena pulau jawa sebagian besar merupakan provinsi sentra beras sementara provinsi yang berada di bagian timur kebanyakan provinsi tidak sentra beras sehingga menimbulkan perbedaan mengenai harga beras.



Gambar 2 : Box Plot Harga Beras

Gambar 2 merupakan diagram kotak garis [box plot] harga beras 32 provinsi Indonesia, terlihat pada gambar 2 kategori provinsi sentra dan tidak sentra beras, menurut kementerian pertanian republik Indonesia terdapat 10 provinsi sentra beras diantaranya ialah provinsi Aceh, Sumatera Barat, Sumatera Utara, Sumatera Selatan, Lampung, Jawa Barat, Jawa Timur, Jawa Tengah, Banten dan Sulawesi Selatan [Kementan, 2019]. Hasil diagram kotak garis harga beras di 32 provinsi Indonesia menunjukkan bahwa harga beras di wilayah sentra beras memiliki nilai harga lebih rendah dibandingkan provinsi sekitarnya yang tidak sentra beras. Provinsi Kalimantan Selatan diduga terdapat data *outlier* karena pada periode tahun 2008-2009 memiliki harga beras yang relatif rendah dibawah Rp 6.000/kg namun pada tahun 2010 hingga 2011 mengalami kenaikan cukup drastis

### 3.2 Pemodelan ARIMA Level Individu

Pemodelan ARIMA menggunakan data latih, diperoleh hasil masing-masing model terbaik per provinsi pada Tabel 2 berdasarkan nilai AICc terkecil dari kandidat model lainnya. Hasil uji diagnostik diperoleh sisaan saling bebas, sisaan berdistribusi normal hanya pada provinsi Lampung sementara pada provinsi lainnya sisaannya tidak berdistribusi normal.

Tabel 2 . Pemodelan ARIMA Level Individu

Provinsi	Model	AICc	P-value	
			L-Jung Box	Jarque-Bera
Aceh	ARIMA[0,1,3]*	1492.10	0.70	0.00
Sumatera Utara	ARIMA[1,1,0][0,0,1]12*	1449.38	0.76	0.00
Sumatera Barat	ARIMA[1,1,1][1,0,0]12*	1604.13	0.56	0.00
Riau	ARIMA[1,1,0]*	1491.94	0.41	0.00
Jambi	ARIMA[1,1,0][1,0,2]12*	1518.62	0.41	0.00
Sumatera Selatan	ARIMA[0,1,1][1,0,1]12*	1541.60	0.58	0.00
Bengkulu	ARIMA[0,1,1][1,0,1]12*	1573.63	0.51	0.00
Lampung	ARIMA[2,1,1][1,1,1]12*	1396.42	0.97	0.70

Provinsi	Model	AICc	P-value	
			L-Jung Box	Jarque-Bera
Bangka Belitung	ARIMA[1,1,1]*	1518.39	0.96	0.00
Kepulauan Riau	ARIMA[1,1,1]*	1600.45	0.06	0.00
Jawa Barat	ARIMA[0,1,1][1,0,1]12*	1529.91	0.61	0.00
Jawa Tengah	ARIMA[0,1,1][0,1,1]12*	1385.05	0.17	0.00
DIY	ARIMA[0,1,1][0,1,1]12*	1387.80	0.18	0.00
Jawa Timur	ARIMA[0,1,1][0,1,1]12*	1356.98	0.09	0.00
Banten	ARIMA[0,1,0][1,1,0]12*	1608.83	0.27	0.00
Bali	ARIMA[1,1,0]*	1548.81	0.15	0.00
NTB	ARIMA[0,1,0][0,1,1]12*	1488.26	0.36	0.00
NTT	ARIMA[0,1,1][1,0,1]12*	1512.90	0.89	0.00
Kalimantan Barat	ARIMA[0,1,0]*	1600.03	1.00	0.00
Kalimantan Tengah	ARIMA[3,1,0]*	1621.85	0.92	0.00
Kalimantan Selatan	ARIMA[0,2,3]*	1530.23	0.75	0.00
Kalimantan Timur	ARIMA[0,1,0]*	1581.55	0.16	0.00
Sulawesi Utara	ARIMA[0,1,3][0,0,2]12*	1587.08	0.82	0.00
Sulawesi Tengah	ARIMA[0,1,1][0,0,2]12*	1545.01	0.65	0.00
Sulawesi Selatan	ARIMA[0,1,0][1,0,0]12*	1519.56	0.45	0.00
Sulawesi Tenggara	ARIMA[4,1,0][0,0,1]12*	1583.92	0.78	0.00
Gorontalo	ARIMA[3,1,0][0,0,1]12*	1613.93	0.55	0.00
Sulawesi Barat	ARIMA[3,1,2][1,0,1]12*	1522.73	0.18	0.00
Maluku	ARIMA[2,1,2]*	1486.20	0.32	0.00
Maluku Utara	ARIMA[1,1,1]*	1570.39	0.98	0.00
Papua Barat	ARIMA[3,1,3]*	1681.32	0.25	0.00
Papua	ARIMA[0,1,1]*	1579.88	0.90	0.00

Keterangan: \*= Model Terbaik

### 3.3 Penggerombolan Data Deret Waktu

Teknik penggerombolan yang digunakan ialah penggerombolan hierarki dengan mencoba semua pautan (*linkage*) dengan ukuran ketidakmiripan yang digunakan adalah jarak korelasi, diperoleh hasil perbandingan beberapa pautan yang disajikan pada Tabel 3.

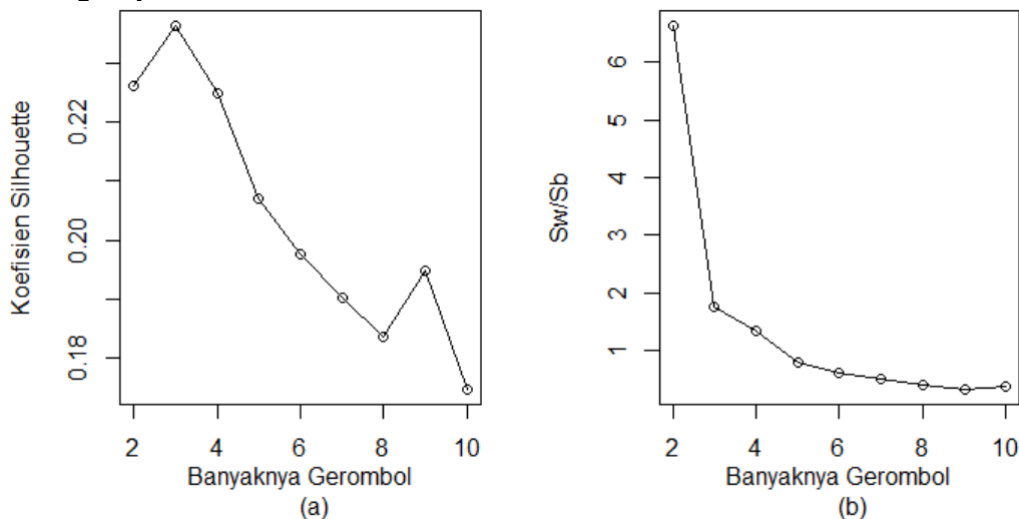
Tabel 3 . Rincian Anggota Gerombol Optimal

Linkage	Korelasi <i>Cophenetic</i>
Average Linkage	0.95
Complete Linkage	0.88
Single Linkage	0.93
Centroid Linkage	0.86

Salah satu ukuran kebaikan penggunaan suatu jarak atau ketidakmiripan dalam penggerombolan data deret waktu adalah koefisien korelasi *cophenetic* yang merupakan korelasi antara jarak *cophenetic* yang didapat dari diagram pohon dengan

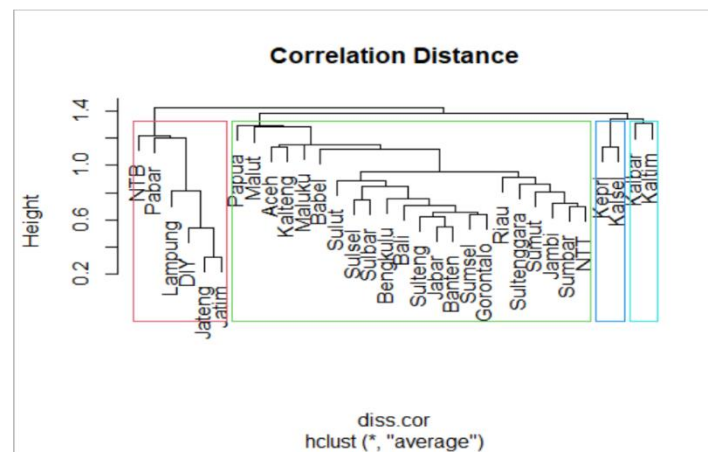


jarak objek semula yang digunakan untuk membuat diagram pohon. Nilai korelasi dalam range  $-1 \leq r \leq 1$ , nilai korelasi *cophenetic* mendekati 1 berarti proses penggerombolan cukup baik. Koefisien korelasi *cophenetic* tertinggi menggunakan *average linkage* sebesar 0.95 artinya proses penggerombolan harga beras berdasarkan nilai ACF untuk masing-masing provinsi cukup baik. Selanjutnya penentuan banyak gerombol optimal dapat diperoleh berdasarkan nilai maksimum koefisien *silhouette* dan nilai minimum dari rasio keragaman [Sw/Sb] pada banyaknya gerombol 2 sampai 10 yang disajikan pada plot pada Gambar 3 menggunakan *average linkage* dengan jarak korelasi.



Gambar 3 : Plot [a] Koefisien silhouette [b] Rasio Keragaman

Salah satu penentuan banyaknya gerombol optimal yang diperoleh dapat melalui interpretasi subyektif koefisien *silhouette* pada setiap kemungkinan jumlah gerombol. Berdasarkan koefisien *silhouette* pada Gambar 3a menunjukkan nilai maksimumnya berada pada banyak gerombol  $k = 3$ . Namun jika gerombol yang dibentuk hanya tiga, maka penyebaran objek pada gerombol belum terlihat karena mengumpul pada suatu gerombol tertentu. Sedangkan untuk kriteria subyektif penggerombolan masuk akal akan didapatkan untuk pembentukan 4 gerombol. Sedangkan rasio keragaman pada gambar 3b menunjukkan semakin banyak gerombol yang digunakan maka rasio keragaman semakin kecil. Sehingga, pemilihan banyak gerombol berdasarkan ukuran rasio keragaman pasti akan memilih banyak gerombol yang paling besar. Namun, jika dilihat kembali pada gambar 3b nilai rasio keragaman antara  $k = 3$  sampai  $k = 10$  tidak terlalu berbeda sehingga banyak gerombol tersebut juga dapat dipertimbangkan sebagai banyak gerombol yang optimal. Banyak gerombol yang optimal berdasarkan koefisien *silhouette* adalah  $k = 4$  sementara berdasarkan rasio keragaman banyak gerombolnya adalah antara  $k = 3$  sampai  $k = 10$ , sementara  $k=4$  memiliki rasio keragaman yang tidak jauh berbeda dengan  $k=3$ , sehingga banyak gerombol optimal yang dipilih adalah  $k = 4$ . sehingga data harga beras di 32 provinsi Indonesia akan digerombolkan sebanyak 4 gerombol. Berikut anggota dari masing-masing gerombol.



Gambar 4 : Dendrogram Hasil Penggerombolan

Tabel 4. Rincian Anggota Gerombol Optimal

Gerombol	Banyaknya Anggota	Anggota
A	6	Nusa Tenggara Barat, Papua Barat, Lampung, D.I.Yogyakarta, Jawa Tengah, Jawa Timur
B	22	Papua, Maluku Utara, Aceh, Kalimantan Tengah, Maluku, Bangka Belitung, Sulawesi Utara, Sulawesi Selatan, Sulawesi Barat, Bengkulu, Bali, Sulawesi Tengah, Jawa Barat, Banten, Sumatera Selatan, Gorontalo, Riau, Sulawesi Tenggara, Sumatera Utara, Jambi, Sumatera Barat, Nusa Tenggara Timur
C	2	Kepulauan Riau, Kalimantan Selatan
D	2	Kalimantan Barat, Kalimantan Timur

Kesesuaian hasil gerombol dengan provinsi sentra dan tidak sentra beras bisa saja berbeda, karena harga beras di lapangan itu tergantung banyak hal, seperti dari produksi, persediaannya, tergantung pada manajemen stok [yang dilakukan oleh bulog atau pihak terkait, selaku pengambilan kebijakan harga beras] dan pola distribusi itu sendiri. Sehingga terkait dengan sentra dan tidak sentra hanya dilihat satu dimensi saja yaitu dari aspek produksi, sehingga wajar saja apabila hasil gerombol yang diperoleh itu bisa berbeda dalam artian tidak selalu bahwa provinsi sentra beras berada dalam satu gerombol yang sama bisa saja ada irisan antar gerombol yang diperoleh berdasarkan provinsi sentra beras, seperti hasil yang diperoleh pada penelitian ini, dengan adanya kecenderungan provinsi sentra dan tidak sentra produksi beras cenderung berada pada Gerombol A dan B, sedangkan wilayah yang termasuk defisit konsumsi beras dan lokasi sekitarnya berjauhan dengan provinsi sentra beras berada pada Gerombol C dan D.

### 3.4 Pemodelan ARIMA Level Gerombol

Pemodelan pada level gerombol memerlukan data perwakilan dari setiap gerombol. Penghitungan wakil dari masing-masing gerombol adalah dengan menghitung nilai rata-rata pada anggota dalam suatu gerombol (*prototype*). Hasil pengujian diagnostik sisaan pada Tabel 5 menunjukkan bahwa, untuk semua model terbaik pada 4 gerombol memiliki hasil yang sama yaitu sisaan saling bebas hal ini di tunjukan dengan nilai *p*-

value pada uji *L-jung Box* > alpha : 0.05. dan sisaannya tidak berdistribusi normal hal ini di tunjukkan dengan nilai *p-value* pada uji *Jarque-Bera* < alpha : 0.05.

Tabel 5 . Pemodelan ARIMA Level Gerombol

Gerombol	Model	AICc	P-value	
			L-Jung Box	Jarque-Bera
A	ARIMA[2,1,2][1,1,1] <sub>12</sub> *	1353.60	0.19	0.00
B	ARIMA[0,1,0][0,1,1] <sub>12</sub> *	1276.50	0.50	0.00
C	ARIMA[0,1,2]*	1521.54	0.09	0.00
D	ARIMA[0,1,0]*	1569.70	0.99	0.00

### 3.5 Perbandingan Pemodelan ARIMA Level Individu dan Gerombol

Perbandingan akan dilakukan secara satu persatu dan menggunakan rata-rata evaluasi prediksi dan ramalan. Rata-rata evaluasi pemodelan dan ramalan ARIMA level individu dan gerombol di ulang sebanyak 3 kali dengan pertimbangan untuk mengetahui konsistensi hasil peramalan terhadap suatu model yang diperoleh. Rata-rata evaluasi prediksi dan ramalan digunakan untuk melihat perbedaan pemodelan dengan penggerombolan dan tanpa penggerombolan. Secara umum sebagaimana ditunjukkan pada tabel 6.

Tabel 6. Perbandingan MAPE

Uji-z	MAPE	
	<i>Training</i>	<i>Testing</i>
Nilai-z	1.17	-1.98
Nilai-p	0.245	0.052

Nilai rata-rata antara MAPE data *training* dan *testing* level individu dan gerombol 3 kali ulangan tersebut tidak berbeda jauh, untuk MAPE data *training* model level individu diperoleh 1.24% sedangkan untuk model level gerombol 1.16%, sementara MAPE untuk data *testing* untuk model level individu diperoleh 3.36% sedangkan untuk model level gerombol 4.27%. Selain itu secara formal dilakukan pula uji z pada Tabel 6 untuk menguji perbedaan antara MAPE pada level individu dengan level gerombol pada data *training* dan *testing* 3 kali ulangan. *P-value* yang dihasilkan sebesar 0.245 untuk data *training*, sedangkan *p-value* data *testing* 0.052, nilai tersebut tidak signifikan pada taraf nyata 5%.

## 4. Simpulan dan Saran

Pada kajian model peramalan data deret waktu dengan peubah banyak dapat dilakukan penggerombolan terlebih dahulu agar lebih efisien. Kedua pendekatan model peramalan ARIMA level individu dan ARIMA level gerombol dilakukan untuk mengetahui sejauh mana performa model yang diperoleh. Pemodelan ARIMA level individu tidak lebih baik dibandingkan pemodelan level gerombol, hal ini dibuktikan dengan 4 gerombol memberikan efisiensi yang tinggi karena hanya melakukan 4 kali pemodelan menghasilkan akurasi tidak jauh berbeda dengan pemodelan level individu. Hasil peramalan harga beras level individu dan gerombol cenderung mengikuti data aktual dan memiliki nilai MAPE yang diperoleh tidak melebihi 10%, sehingga pemodelan level gerombol tergolong baik dan representatif untuk dilakukan

peramalan.

Dalam penelitian ini penggerombolan data deret waktu menggunakan pendekatan berbasis karakteristik data (*eatured-based*) dengan jarak korelasi, diperoleh hasil penggunaan jarak korelasi hanya menangkap kemiripan pola data deret waktu, namun belum memperhatikan jarak atau nilai dari suatu data. Sehingga penelitian selanjutnya dapat dilakukan pengkajian jarak atau melakukan kombinasi jarak yang mampu menangkap pola dan nilai dari suatu data deret waktu.

## Daftar Pustaka

- [BPS] Badan Pusat Statistik. (2018). *Statistik Harga konsumen Perdesaan Kelompok Makanan*. Jakarta [ID]: Badan Pusat Statistik.
- Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis: with applications in R* (Vol. 2). Springer.
- D'Urso, P., & Maharaj, E. A. (2009). Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, 160(24): 3565–3589.
- Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., & Boesiger, P. (1998). A new correlation-based fuzzy logic clustering algorithm for FMRI. *Magnetic Resonance in Medicine*, 40(2): 249–260.
- Kalkstein, L. S., Tan, G., and Skindlov, J. A. (1987). An evaluation of three clustering procedures for use in synoptic climatological classification. *Journal of Applied Meteorology and Climatology*, 26(6): 717–730.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- [Kementan RI] Kementerian Pertanian Republik Indonesia. 2019. 10 Besar Provinsi Penghasil Beras [Internet]. Retrieved from <https://www.pertanian.go.id/home/?show=news&act=view&id=4425>
- Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. Springer Science & Business Media.
- Maharaj, E. A., dan Inder, B. A. (1999). *Forecasting time series from clusters*.
- Kumar M, N. R. P. (2008). *Combining forecasting using clustering*.
- Mardianto, I., Gunawan, M. I., Sugiarto, D., Rochman, A., & others. (2020). Comparison of rice price forecasting using the ARIMA method on amazon forecast and sagemaker. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(3): 537–543.
- Montero, P., dan Vilar, J. A. (2015). TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62: 1–43.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Munthe, A. D. (2019). Penerapan clustering time series untuk menggerombolkan provinsi di Indonesia berdasarkan nilai produksi padi. *Jurnal Litbang Sukowati: Media Penelitian Dan Pengembangan*, 2(2): 11–11.
- Novidianto, R., dan Dani Andrea Tri, R. (2020). Analisis klaster kasus aktif COVID-19 menurut provinsi di Indonesia berdasarkan data deret waktu. *Jurnal Aplikasi Statistika dan Komputasi Statistik*, 15–24.
- Rencher, A. C. (2005). *A review of "Methods of Multivariate Analysis, ."* Taylor & Francis.
- Siswanto, E., Sinaga, B. M., others. (2018). Dampak Kebijakan Perberasan Pada Pasar Beras dan Kesejahteraan Produsen dan Konsumen Beras di Indonesia. *Jurnal Ilmu Pertanian Indonesia*, 23(2): 93–100.

- Sokal, R. R., dan Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 33–40.
- Utami, B. (n.d.). *Model VARX untuk Peramalan Inflasi Menurut Sub Kelompok Komoditi di Jakarta dengan Pendekatan TSClust sebagai Preprocessing* (PhD Thesis). Bogor Agricultural University (IPB).
- Wei. (2006). *Time Series Analysis, Univariate and Multivariate Methods* (2nd ed.). New York [US]: Pearson Education, Inc.
- Wijaya, S. U., & Ngatini, N. N. (2020). Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series. *Limits: Journal of Mathematics and Its Applications*, 17(1): 51–66.