

# ALTERNATIF PENGGEROMBOLAN DATA DERET WAKTU DENGAN KONDISI TERDAPAT DATA KOSONG\*

(Studi Kasus Penggerombolan Provinsi di Indonesia Berdasarkan  
Data Deret Waktu Rasio Gini tahun 2007 – 2017)

Yusma Yanti<sup>1</sup>, Septian Rahardiantoro<sup>2‡</sup>

<sup>1</sup> Department of Computer Sciences, Pakuan University, Indonesia, yusma.yanti@unpak.ac.id

<sup>2</sup>Department of Statistics, Bogor Agricultural University (IPB), Indonesia,  
rahardiantoro\_14@apps.ipb.ac.id

‡corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**

**Vol 2 No 1 (2018), 13 - 22**

Copyright © 2018 Yusma Yanti, and Septian Rahardiantoro. 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

Panel data describes a condition in which there are many observations with each observation observed periodically over a period of time. The observation clustering context based on this data is known as Clustering of Time Series Data. Many methods are developed based on fluctuating time series data conditions. However, missing data causes problems in this analysis. Missing data is the unavailability of data value on an observation because there is no information related to it. This study attempts to provide an alternative method of clustering observations on data with time series containing missing data by utilizing correlation matrices converted into Euclid distance matrices which are subsequently applied by the hierarchical clustering method. The simulation process was done to see the goodness of alternative method with common method used in data with 0%, 10%, 20% and 40% missing data condition. The result was obtained that the accuracy of the observation bundling on the proposed alternative method is always better than the commonly used method. Furthermore, the implementation was done on the annual gini ratio data of each province in Indonesia in 2007 to 2017 which contained missing data in North Kalimantan Province. There were 2 clusters of province with different characteristics.

**Keywords:** clustering time series, correlation matrix, euclid distance, gini ratio.

---

\* Received Des 2017; Accepted Apr 2018; Published online on Apr 2018

## 1. Pendahuluan

Secara umum terdapat tiga jenis data yang seringkali ditemui, yaitu data *cross section*, data *time series*, dan data panel. Perbedaan pada masing-masing jenis data tersebut ialah pada banyak observasinya dan ada tidaknya efek waktu di dalamnya. Data *cross section* memiliki banyak observasi yang diamati pada satu waktu, data *time series* memiliki satu objek yang diamati secara berkala pada suatu periode waktu tertentu, sedangkan data panel memiliki banyak observasi yang diteliti yang diamati berkala pada suatu periode waktu tertentu (Griliches & Intriligator, 1986). Penelitian ini dilakukan untuk persoalan penggerombolan observasi pada data panel.

Penggerombolan observasi pada data panel seringkali disebut dengan istilah *Clustering Time Series*. Inti dari analisis ini ialah dalam penentuan jarak (*dissimilarity*) yang tepat digunakan pada data panel (Montero & Vilar, 2014). Selanjutnya ketika jarak sudah ditentukan maka metode penggerombolan dapat dilakukan. Namun, seringkali pada data terdapat data kosong yaitu tidak tersedianya dengan lengkap informasi pada data karena data pada observasi tersebut tidak dapat dilakukan proses pengukurannya. Adanya data kosong pada data panel diduga dapat menurunkan akurasi penggerombolan yang dilakukan. Untuk itu, penelitian ini dilakukan untuk mengkaji kinerja metode *clustering time series* ketika terdapat data kosong.

Selain itu, penelitian ini juga mencoba memberikan metode alternatif pada *clustering time series* ketika terdapat data kosong dengan mengkonversi matriks data menjadi matriks korelasi yang selanjutnya dikonversi menjadi matriks jarak Euclid. Pada matriks jarak yang terbentuk, digunakan untuk penggerombolan observasinya dengan metode penggerombolan berhierarki. Kajian dilakukan dengan simulasi pada data yang dibangkitkan yang memiliki 3 buah gerombol, dengan tanpa adanya data kosong, terdapat 10% data kosong, 20% data kosong, dan 40% data kosong. Metode alternatif ini juga sangat baik digunakan untuk penggerombolan BPRS berdasarkan pertumbuhan nilai bisnisnya yang melibatkan banyak peubah (Sartono *et al.*, 2017). Selanjutnya, metode alternatif tersebut diaplikasikan pada data rasio gini Indonesia yang merupakan salah satu alat yang digunakan untuk mengukur pemerataan penduduk di Indonesia. Nilai rasio gini diukur setiap tahun pada tahun 2007 sampai 2017 pada setiap provinsi di Indonesia. Data kosong terdapat pada data tersebut yaitu pada provinsi Kalimantan Utara yang nilai rasio gininya baru ada sejak tahun 2015. Pemetaan ini dapat memberikan gambaran provinsi dengan tingkat pemerataan yang baik atau tingkat ketimpangan yang sempurna. Jadi diharapkan pemerintah dapat memberikan tindak lanjut yang tepat pada provinsi yang ada. Keseluruhan penerapan metode penelitian ini menggunakan software R.

Ringkasnya, penelitian ini bertujuan untuk mengkaji kebaikan metode *clustering time series* untuk penggerombolan observasi pada data yang mengandung data kosong. Selain itu, penelitian ini juga mengkaji kebaikan metode alternatif *clustering time series* dengan memanfaatkan matriks korelasi data yang dikonversi menjadi matriks jarak euclid pada data dengan mengandung data kosong. Serta aplikasi metode tersebut untuk mengelompokkan provinsi di Indonesia berdasarkan rasio gini setiap provinsi yang terdapat data kosong.

## 2. Rasio Gini

Rasio gini merupakan salah satu alat yang digunakan untuk mengukur ketidakmerataan sebaran penduduk. Berdasarkan kurva Lorenz yaitu sebuah kurva tebaran kumulatif yang membandingkan sebaran dari suatu variabel tertentu (misalnya pendapatan) dengan sebaran uniform (seragam) yang mewakili persentase kumulatif penduduk. Rasio Gini bernilai antara 0 sampai 1, dengan nilai 0 menunjukkan adanya pemerataan pendapatan yang sempurna, atau setiap orang memiliki pendapatan yang sama. Sebaliknya, nilai 1 menunjukkan bahwa terdapat ketimpangan pendapatan. Jadi semakin besar nilai rasio gini, semakin besar ketidakmerataan antara sebaran penduduk dan jumlah lokasi. Rumus dari rasio gini adalah sebagai berikut (BPS, 2017):

$$GR = 1 - \sum_{i=1}^n P_i (F_i + F_{i-1})$$

dengan:

$GR$  : Koefisien rasio gini

$P_i$  : Frekuensi penduduk dalam kelas pengeluaran ke- $i$

$F_i$  : Frekuensi kumulatif dari total pengeluaran dalam kelas pengeluaran ke- $i$

$F_{i-1}$  : Frekuensi kumulatif dari total pengeluaran dalam kelas pengeluaran ke- $(i - 1)$

## 3. Analisis Gerombol

Analisis gerombol merupakan teknik analisis yang memanfaatkan variabel data untuk mengelompokkan objek-objek dengan karakteristik yang serupa. Pada analisis gerombol tidak memiliki variabel respon, sehingga data yang ada hanya memiliki variabel-variabel indikator yang akan digunakan untuk mengelompokkan objek yang ada. Objek dengan kesamaan karakteristik yang dimiliki cenderung dimasukkan ke dalam satu kelompok dan sebaliknya yang memiliki karakteristik yang berbeda pada kelompok yang lain (Mattjik *et al.*, 2011).

Salah satu penerapan dari analisis gerombol adalah untuk pemetaan objek-objek dalam kelompok yang sejenis. Akibatnya dalam kasus pada penelitian ini dikaji pengelompokan provinsi dengan karakter yang sama beserta karakteristik yang berbeda untuk antar kelompok yang berbeda. Sehingga dari hasil pengelompokan yang diperoleh diharapkan dapat membantu pemerintah untuk memberikan perlakuan atau tindakan yang tepat pada provinsi yang ada.

Analisis gerombol terdapat dua jenis, yaitu analisis gerombol berhierarki dan analisis gerombol non-hierarki. Metode analisis gerombol berhierarki merupakan salah satu cara analisis gerombol di mana belum diketahui kemiripan sifat dan jumlah gerombol yang diinginkan. Terdapat dua cara pada metode berhierarki ini, yaitu penggabungan (*aglomerative*) dan pemisalahan (*devisive*). Metode hierarki *aglomerative* melakukan penggabungan objek secara bertahap sehingga dihasilkan satu gerombol. Sebaliknya pada metode hierarki *devisive* melakukan pemisahan bertahap dari sebuah gerombol yang besar menjadi gerombol kecil hingga satu gerombol hanya terdiri dari satu objek (Johnson & Wichern, 2007).

Salah satu konsep jarak yang sering digunakan untuk penentuan objek dengan kesamaan karakteristik adalah jarak Euclid. Hasil dari jarak Euclid menginterpretasikan bahwa objek dengan jarak yang jauh memiliki karakteristik yang berbeda. Sedangkan objek dengan jarak Euclid yang dekat memiliki karakteristik yang

sama (Mattjik & Sumertajaya, 2011) (Johnson & Wichern, 2007). Misalkan diketahui vektor observasi  $x_a$  dan  $x_b$ ;  $a \neq b$ , rumus untuk mencari jarak euclid antar observasi vektor observasi tersebut ialah:

$$d(x_a, x_b) = \sqrt{(x_a - x_b)'(x_a - x_b)}$$

Penelitian ini menggunakan analisis gerombol dengan metode berhierarki *agglomerative*. Metode *agglomerative* yang digunakan ialah metode Ward. Metode ini meminimumkan jumlah kuadrat pada keseluruhan data pada observasi untuk gerombol yang sama. Kemudian, tahapan dilanjutkan dengan memasukkan observasi ke dalam gerombol yang sama sehingga jumlah kuadrat galat yang ada di dalam gerombol bernilai lebih kecil. Berikut ini merupakan algoritma metode Ward yang digunakan (Ward Jr, 1963).

1. Menghitung matriks jarak kuadrat euclid dari setiap pasang vektor observasi  $x_a$  dan  $x_b$ ;  $a \neq b$
2. Menghitung jumlah kuadrat galat yang diperoleh dari

$$JKG = \sum_{l=1}^s (d^2(x_a, x_b))_l^2 - \frac{1}{s} \left( \sum_{l=1}^s d^2(x_a, x_b)_l \right)^2$$

dengan  $d^2(x_a, x_b)$  menyatakan jarak kuadrat euclid pasangan vektor observasi  $x_a$  dan  $x_b$ ;  $a \neq b$  untuk pasangan ke  $l = 1, 2, \dots, s$ . Pada awalnya masing-masing observasi menjadi satu gerombol sehingga nilai  $JKG$  gerombol awal akan bernilai 0.

3. Penggabungan anggota gerombol dilakukan berdasarkan nilai jarak kuadrat euclid terkecil. Proses ini dilakukan sampai diperoleh nilai  $JKG$  total yang minimum.

#### 4. Analisis *Cluster Time Series*

Penggunaan analisis gerombol biasanya pada data yang bersifat statis. Namun pada data rasio gini Indonesia yang dipengaruhi oleh waktu menyebabkan metode ini tidak dapat digunakan. Metode analisis *cluster time series* adalah salah satu solusi yang dapat digunakan. Pada metode ini, algoritma yang digunakan pada data statis akan dimodifikasi sedemikian rupa sehingga bisa digunakan pada data deret waktu. Akibatnya algoritma yang ada untuk pengelompokan data statis bisa langsung digunakan pada data deret waktu, dengan modifikasi utama pada penggantian jarak/kemiripan pada data statis yang disesuaikan untuk data deret waktu.

Menurut Liao (2005) terdapat tiga kategori analisis *cluster times series*. Pertama pendekatan berbasis data mentah, cara kerjanya dengan menghitung jarak antar series menggunakan data mentah. Hasil yang diperoleh digunakan untuk proses penggerombolan. Jarak yang sering digunakan adalah jarak Euclid atau jarak yang didasarkan pada nilai korelasi. Kedua, pendekatan dengan menyarikan data mentah dengan cara menghilangkan pengaruh data *noise* dan mengurangi dimensi data, setelahnya baru dilakukan perhitungan jarak dan dilakukan proses penggerombolan. Ketiga, pendekatan dengan model yang telah dibentuk dari data mentah. Sehingga nantinya penggerombolan diduga menggunakan model yang ada. Penelitian ini menggunakan metode alternatif pertama dengan memanfaatkan matriks korelasi data mentah untuk proses awal penentuan matriks jarak dalam proses penggerombolan.

## 5. Metodologi Penelitian

### 5.1 Data

Terdapat dua jenis data yang digunakan pada penelitian ini, yaitu data simulasi dan data deret waktu rasio gini seluruh provinsi di Indonesia pada tahun 2007 sampai 2017. Data simulasi dibangkitkan dengan sebanyak 300 observasi yang terbagi menjadi tiga gerombol berdasarkan pola deret waktunya. Masing-masing gerombol memiliki 100 observasi dengan 50 periode waktu, yang dibangkitkan dengan kriteria: gerombol 1 mengikuti ARIMA(1,1,0) dengan penambahan unsur tren meningkat; gerombol 2 mengikuti pola konstan (stasioner); dan gerombol 3 mengikuti ARIMA(0,1,1) dengan penambahan unsur tren menurun. Pada data yang terbentuk dikaji dengan 4 kriteria, yaitu tanpa data kosong, terdapat data kosong 10%, terdapat data kosong 20%, dan terdapat data kosong 40%. Kajian pada data simulasi dilakukan dengan metode penggerombolan *agglomerative* dengan metode Ward baik secara langsung maupun dengan proses pembentukan matriks korelasi pada tahapan awalnya. Semua proses kajian tersebut dilakukan dengan ulangan sebanyak 1000 kali.

Data penerapan yang digunakan pada penelitian ini adalah data rasio gini Indonesia tahun 2007 sampai 2017 yang dapat diakses melalui situs <https://www.bps.go.id>. Pada data ini terdapat 34 provinsi yang nantinya yang akan menjadi objek penelitian dengan 11 periode waktu. Nilai rasio gini yang tersedia hampir di setiap tahun dari tahun 1996. Akan tetapi dengan pemekaran provinsi yang dilakukan pemerintah sehingga terbentuk provinsi baru yang menyebabkan timbulnya data kosong. Salah satunya ialah pada provinsi Kalimantan Utara yang memiliki nilai rasio gini mulai tahun 2015.

### 5.2 Metode

Pada tahapan kajian simulasi akan diterapkan dua metode penggerombolan, yaitu metode *agglomerative* Ward dengan unsur waktu diabaikan polanya, dan metode *agglomerative* Ward dengan memperhatikan pola deret waktu setiap observasinya. Metode 1 tidak lain adalah metode penggerombolan yang secara umum digunakan. Di sisi lain, metode 2 merupakan alternatif metode yang dikaji pada penelitian ini. Pola deret waktu pada setiap observasi diidentifikasi dengan menerapkan keeratn pola hubungan antar observasinya, sehingga matriks korelasi dipilih untuk dapat menjelaskan pola yang terbentuk. Lebih lanjut, berikut ini tahapan metode yang diterapkan pada data simulasi yang telah terbentuk.

1. Pembangkitan data simulasi dengan 3 gerombol masing-masing 100 observasi untuk  $t = 1, 2, \dots, 50$ :

- a. gerombol 1 memiliki pola tren naik dengan persamaan observasi ke- $i$ :

$$x_{t,i} = \mu + x_{t-1,i} + \phi_1(x_{t-1,i} - x_{t-2,i}) + r_{1,i}$$

dengan  $\phi_1 = 0.7$  dan  $r_{1,i} = t$

- b. gerombol 2 memiliki pola stasioner

- c. gerombol 3 memiliki pola tren turun dengan persamaan observasi ke- $i$ :

$$x_{t,i} = x_{t-1,i} - \theta_1 e_{t-1,i} + r_{2,i}$$

dengan  $\theta_1 = 0.5$  dan  $r_{2,i} = 50 + 1 - t$

2. Pada data simulasi yang telah dibangkitkan, selanjutnya dibentuk berdasarkan 4 kriteria yaitu tanpa data kosong, terdapat data kosong 10%, 20%, dan 40%. Proses penyisihan amatan menjadi data kosong dilakukan secara acak dengan amatan yang berurutan periode waktunya.
3. Menggerombolkan data simulasi yang telah dibangkitkan dengan metode 1 (metode *agglomerative ward*) dengan dibentuk gerombol sebanyak 3 buah.
4. Menggerombolkan data simulasi tersebut dengan metode 2 yang mengikuti algoritma sebagai berikut:

- a. membentuk matriks korelasi antar observasi  $C_{n \times n} = [c_{ij}]$ ;  $c_{ij} = 1, i = j$
- b. mengkonversi matriks korelasi  $C_{n \times n}$  menjadi matriks jarak euclid  $D_{n \times n} = [d_{ij}]$ , dengan (Alfakih & Wolkowicz, 2002)

$$d_{ij} = \left( \frac{1 - c_{ij}}{1 + c_{ij}} \right)^\beta ; \beta > 0$$

Pada penelitian ini, nilai  $\beta$  ditetapkan dengan  $\beta = 0.5$ .

- c. Memanfaatkan matriks jarak euclid untuk penggerombolan dengan metode *agglomerative Ward* dengan dibentuk gerombol sebanyak 3 buah.
5. Mengevaluasi kebaikan penggerombolan pada metode 1 dan metode 2 dengan kriteria evaluasi penggerombolan sebagai berikut (Montero & Vilar, 2014)

$$Ev(T, P) = \frac{1}{k} \sum_{p=1}^k \max_{1 \leq q \leq k} Ev(T_p, P_q)$$

dengan

$$Ev(T_p, P_q) = \frac{2|T_p \cap P_q|}{|T_p| + |P_q|}$$

dengan  $T$  merupakan gerombol asal observasi,  $P$  merupakan gerombol prediksi observasi, serta nilai  $k$  adalah banyaknya gerombol terbentuk. Semakin besar nilai  $Ev(T, P)$  mengindikasikan semakin akurat proses penggerombolan yang dilakukan.

6. Ulangi tahapan 1 sampai 5 sebanyak 1000 kali.

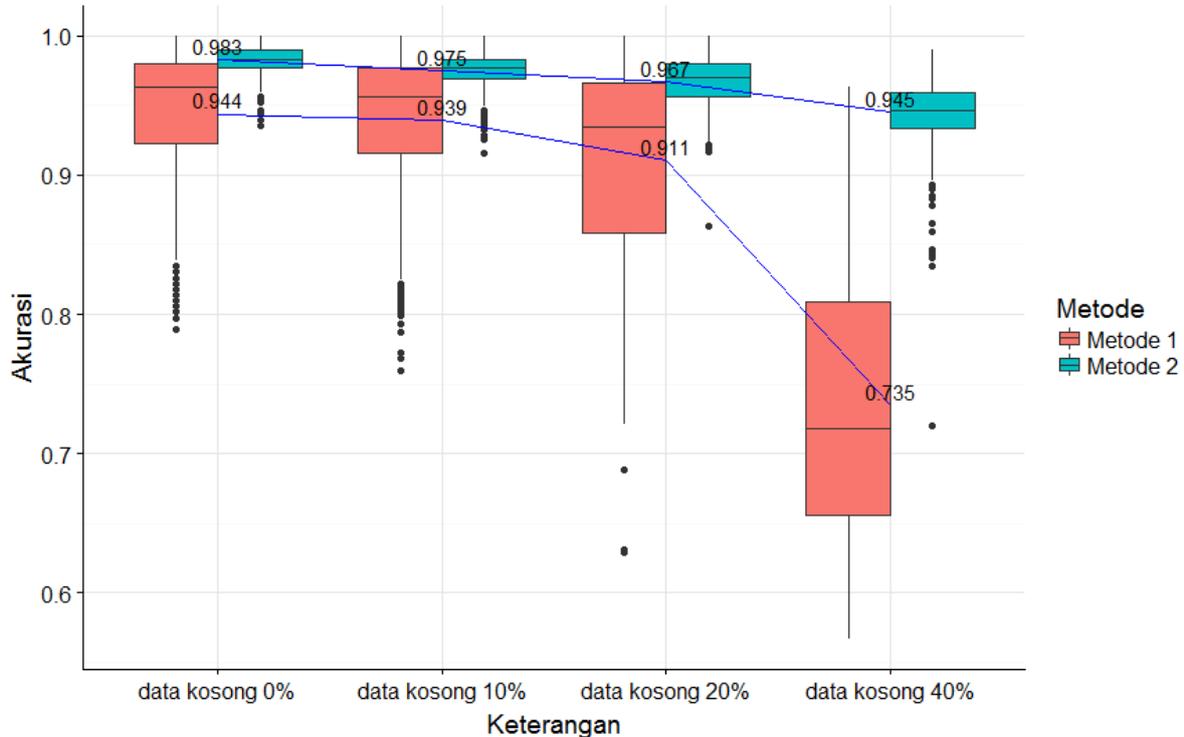
Selanjutnya, data rasio gini seluruh provinsi di Indonesia tahun 2007 sampai 2017 digunakan sebagai sajian kasus. Data ini terdapat 34 observasi dengan 11 periode waktu tahunan. Metode 2 diaplikasikan pada data tersebut dengan pembentukan kelompok berdasarkan banyak kelompok optimum yang terbentuk. Ukuran kesamaan pola sebaran pertumbuhan penduduk setiap provinsi diukur dengan korelasi antar provinsi. Dua provinsi dengan pola sebaran pertumbuhan penduduk yang sama akan memiliki korelasi yang positif dan bernilai mendekati 1, sedangkan dua provinsi yang pola pertumbuhannya berlawanan akan cenderung bernilai korelasi negatif. Nilai korelasi inilah yang dijadikan dasar penentuan gerombol dimana provinsi dengan korelasi yang besar akan berada pada gerombol yang sama.

## 6. Hasil dan Pembahasan

### 6.1 Kajian Simulasi

Simulasi dilakukan dengan 1000 kali ulangan pada 4 kriteria, yaitu data tanpa data kosong, terdapat data kosong 10%, 20%, dan 40%. Gambar 1 menunjukkan diagram

kotak garis nilai akurasi penggerombolan  $Ev(T, P)$  pada ke-empat kriteria tersebut pada metode 1 dan metode 2. Secara umum pada metode 1 dan metode 2 yang diterapkan, semakin banyak data kosong menyebabkan nilai evaluasi penggerombolannya menurun. Hal ini mengindikasikan adanya data kosong pada data akan menurunkan akurasi penggerombolannya.



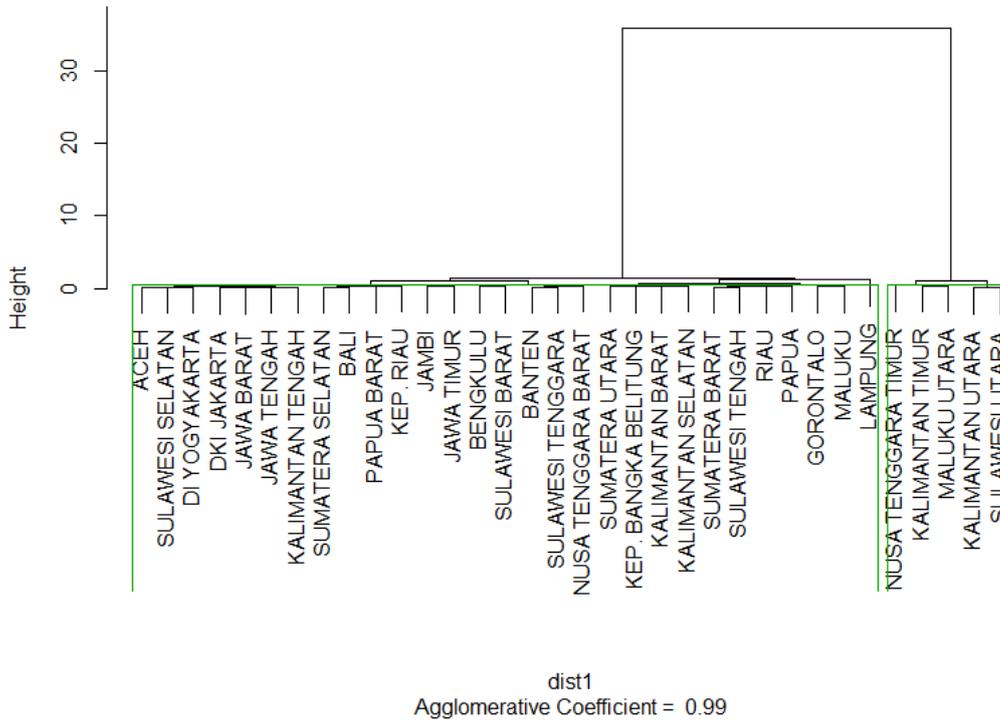
Gambar 1: Diagram kotak garis nilai akurasi evaluasi penggerombolan pada metode 1 dan metode 2 untuk 4 kriteria data.

Meskipun demikian, terlihat bahwa metode 2 yang diusulkan lebih efisien dalam proses penggerombolan pada data yang terdapat data kosong dibandingkan dengan metode 1. Nilai evaluasi penggerombolan pada metode 2 untuk semua kriteria yang dikaji masih lebih besar dari 0.94. Hal ini mengindikasikan metode 2 masih sangat baik dalam menggerombolkan observasi meskipun terdapat data kosong hingga 40% bagian dari data. Akibatnya, berdasarkan kajian simulasi ini metode 2 yang diusulkan dapat digunakan sebagai alternatif metode yang sangat baik untuk proses penggerombolan data deret waktu dengan terdapat data kosong hingga 40%.

## 6.2 Studi Kasus Penggerombolan Provinsi di Indonesia Berdasarkan Data Deret Waktu Rasio Gini tahun 2007 – 2017

Pada bagian ini akan disajikan proses penggerombolan provinsi di Indonesia berdasarkan nilai rasio gini-nya pada periode 2007 – 2017. Data tersebut terdapat data kosong pada Provinsi Kalimantan Utara yaitu pada tahun 2007 sampai 2014, yang dikarenakan Provinsi Kalimantan Utara termasuk provinsi pemekaran yang relatif baru terbentuk. Oleh karena itu, pada data tersebut terdapat data kosong sebanyak 2.14%. Selanjutnya proses penggerombolan dilakukan dengan menggunakan metode 2 berdasarkan algoritma yang tertera pada bagian metode poin

ke-4. Banyaknya gerombol yang terbentuk tidak semata-mata langsung ditentukan, namun berdasarkan banyaknya gerombol optimum dari dendogram yang terbentuk.



Gambar 2: Dendogram penggerombolan provinsi di Indonesia berdasarkan nilai rasio gini tahun 2007 – 2017 dengan metode 2 yang diusulkan



Gambar 3: Peta provinsi di Indonesia berdasarkan gerombol yang terbentuk

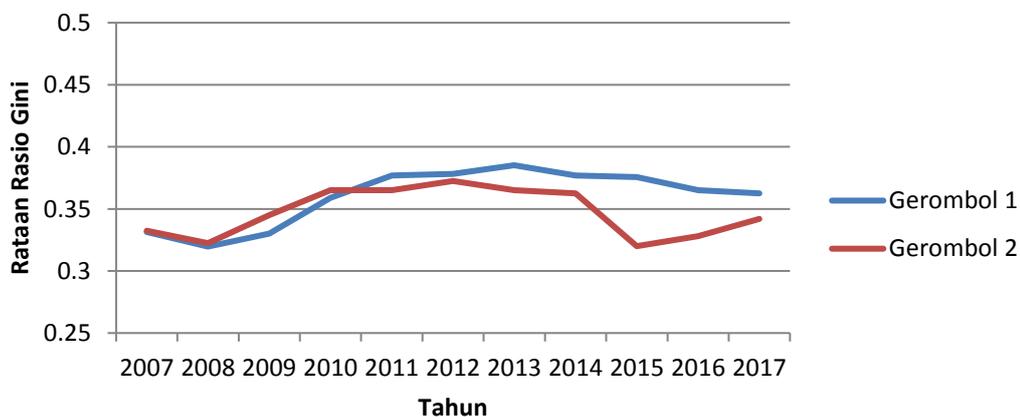
Gambar 2 merupakan hasil dendogram pada proses penggerombolan dengan metode 2 yang diterapkan. Pada dendogram tersebut dapat ditentukan banyaknya gerombol optimum yaitu sebanyak 2 buah. Gerombol pertama terdiri dari 29 provinsi dan gerombol kedua terdiri dari 5 provinsi. Gerombol pertama dengan 29 provinsi ini terdiri dari 10 Provinsi di pulau Sumatera, 6 provinsi di pulau Jawa, 5 provinsi di pulau Sulawesi, 3 provinsi di pulau Kalimantan, 2 provinsi di pulau Nusa Dua dan Bali, serta

3 provinsi dari pulau Maluku dan Papua. Gerombol pertama cenderung terdiri dari provinsi yang sudah lama terbentuk. Gerombol kedua terdiri dari provinsi NTT, Kalimantan Timur, Kalimantan Utara, Maluku Utara, dan Sulawesi utara. Dari lima provinsi tersebut hanya NTT yang bukan merupakan provinsi baru, sedangkan empat provinsi lainnya merupakan provinsi baru yang terbentuk dari hasil pemekaran wilayah. Hasil gerombol terbentuk juga disajikan dalam bentuk peta provinsi di Indonesia pada Gambar 3.

Tabel 1 menyatakan nilai rata-rasio gini pada gerombol 1 dan gerombol 2 yang terbentuk. Rataan rasio gini untuk gerombol 1 bernilai 0.360 yang lebih besar daripada gerombol 2 yang sebesar 0.339, hal ini mengindikasikan bahwa pendapatan penduduk pada provinsi-provinsi di gerombol 1 relatif tidak merata atau timpang. Selain itu, Gambar 4 menyatakan pola deret waktu dari rata-rasio gini untuk setiap gerombol. Pola pada gerombol 1, secara rata-rata relatif meningkat sampai tahun 2013, dan mulai menurun setelahnya. Hal ini mengindikasikan bahwa pemerataan pendapatan penduduk secara agregat pada pulau-pulau dominan di Indonesia mulai membaik. Pola nilai rata-rasio gini pada gerombol 2 relatif lebih rendah daripada gerombol 1, namun dari tahun 2015 nilai rata-rasio gini-nya mulai meningkat. Hal ini menandakan bahwa pada provinsi-provinsi di gerombol 2 pemerataan pendapatan penduduk relatif kurang merata.

Tabel 1: Rataan nilai rasio gini untuk masing-masing gerombol terbentuk.

Gerombol	Rataan nilai rasio gini
1	0.360
2	0.339



Gambar 4: Diagram garis pola deret waktu tahunan rata-rasio gini pada gerombol terbentuk

## 7. Simpulan

Pada kajian simulasi yang dilakukan, metode penggerombolan data deret waktu dengan memanfaatkan matriks korelasi observasinya dapat dijadikan alternatif solusi metode yang sangat baik ketika pada data terdapat data kosong. Hal ini diindikasikan dari nilai akurasi yang dihasilkan masih lebih besar dari 0.94 meskipun terdapat data

kosong pada data mencapai 40%. Selanjutnya, kajian kasus penggerombolan provinsi di Indonesia berdasarkan rasio gini tahun 2007 – 2017, menghasilkan 2 gerombol optimum. Gerombol yang terbentuk diindikasikan dari perbedaan karakteristik provinsi dari pemekaran wilayah, dengan provinsi yang baru terbentuk memiliki nilai rata-ran rasio gini yang lebih kecil. Hal ini menandakan bahwa provinsi-provinsi yang relatif baru, memiliki kondisi pemerataan pendapatan penduduk yang relatif baik dibandingkan provinsi-provinsi lainnya.

### Daftar Pustaka

- Alfakih, A. Y., & Wolkowicz, H. (2002). Two theorems on Euclidean distance matrices and Gale transform. *Linear Algebra and its Applications*, 340(1-3), 149-154.
- BPS. (2017, Des 17). Koefisien Gini. Diambil dari <https://sirusa.bps.go.id/index.php?r=indikator/view&id=22>.
- Griliches, Z., & Intriligator, M. D. (1983). *Handbook of Econometrics*, Volume III. Elsevier Science Publishers BV.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th Edition. Pearson.
- Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, 38(11), 1857-1874.
- Mattjik, A. A., Sumertajaya, I., Wibawa, G. N. A., & Hadi, A. F. (2011). Sidik peubah ganda dengan menggunakan SAS. Departemen Statistika, IPB.
- Montero, P., & Vilar, J. A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1), 1-43.
- Sartono, B., Rahardiantoro, S., Suhaeni, C., Irianto, E. A., & Maulidya, U. (2017, March). Segmentation of Sharia Rural Banks based on Growth of the Business Performance. In *IOP Conference Series: Earth and Environmental Science* (Vol. 58, No. 1, p. 012005). IOP Publishing.
- Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), 236-244.