

PENGEMBANGAN ANALISIS GEROMBOL BERHIRARKI DENGAN KETERGANTUNGAN SPASIAL PADA INDIKATOR MAKRO SOSIAL EKONOMI DI KABUPATEN/KOTA PROVINSI SULAWESI TENGAH*

Iman Setiawan^{1‡}, Nur'eni², and Sritasarwati Putran³

¹Statistics Studies Program, Tadulako University (Untad), Indonesia, i.setiawan@untad.ac.id

²Statistics Studies Program, Tadulako University (Untad), Indonesia, eniocy@yahoo.com

³Statistics Studies Program, Tadulako University (Untad), Indonesia, csrtp19@gmail.com

[‡]corresponding author

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

Vol 4 No 1 (2020), 179 - 188

Copyright © 2020 Iman Setiawan, Nur'eni, and Sritasarwati Putran. 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

This paper develops how the hierarchical clustering analysis uses multivariate variables with spatial dependence on macro social-economic indicator data in Regency/City Central Sulawesi Province. Macro social-economic indicator data used in this paper are the number of criminal cases, per capita expenditure, population density, and Human Development Index of Regency/City of Central Sulawesi Province in 2018. To answer this question, Macro social-economic indicator data was reduced to a new variable using principal component analysis. The new variable was used to identify spatial dependency using the Moran index test. Spatial weight, that meets the Moran index test on the alternative hypothesis (there is a spatial dependency between locations), was used as the spatial dependency distance. Cluster analysis using two distance including variable and spatial dependency distance. The results showed that neighboring Regency/City are in the same cluster (spatial dependency occasion). So that there are five clusters Regency/City in Central Sulawesi Province.

Keywords: hierarchical clustering, macro social economic indicator, principal component analysis, spatial dependency distance.

* Received Nov 2019; Accepted Feb 2020; Published online on Feb 2020

1. Pendahuluan

Data peubah ganda merupakan data dengan peubah yang cukup banyak. Data peubah ganda dapat dengan mudah ditemukan dalam kehidupan sehari-hari salah satunya adalah data indikator makro sosial ekonomi yang tersusun dari data kependudukan, indeks pembangunan manusia, pendapatan perkapita dan lain-lain. Salah satu analisis yang digunakan pada data peubah ganda adalah analisis gerombol. Analisis gerombol merupakan analisis yang bertujuan untuk mencari pola pengelompokan dari suatu objek pada data peubah ganda.

Analisis gerombol terbagi dua yaitu berhirarki dan tak berhirarki. Analisis gerombol berhirarki mempunyai karakteristik jumlah gerombol (k) yang tidak ditentukan terlebih dahulu sebaliknya gerombol tak berhirarki jumlah k perlu ditentukan terlebih dahulu. Misalnya, pada data indikator makro sosial ekonomi belum ada informasi tentang jumlah k sehingga lebih cocok menggunakan analisis gerombol berhirarki. Kemiripan antar objek pada analisis gerombol berhirarki diukur menggunakan jarak antar objek (Johnson & Wichern, 2007).

Chavent *et al.* (2018) mengembangkan analisis gerombol berhirarki dengan menggunakan kendala spasial. Tujuannya adalah menghasilkan gerombol objek berdasarkan peubah ganda pada data dengan tetap memperhatikan kedekatan geografis. Analisis gerombol berhirarki dengan menggunakan kendala spasial menggunakan dua jarak sebagai inputannya. Jarak yang pertama adalah jarak yang merepresentasikan peubah ganda pada data dan jarak yang kedua adalah jarak yang merepresentasikan kedekatan geografis antar objek.

Anselin (1995) menjelaskan bahwa suatu peubah respon dapat dipengaruhi oleh kedekatan geografis objek disekitarnya yang biasa disebut dengan ketergantungan spasial. Jarak yang merepresentasikan kedekatan geografis yang dimaksud oleh Chavent *et al.* (2018) beragam jenisnya dan belum tentu dapat menunjukkan adanya hubungan ketergantungan spasial antara objek terhadap peubah ganda. Salah satu cara mengukur ketergantungan spasial adalah dengan melakukan uji ketergantungan spasial misalnya uji indeks moran seperti yang dilakukan Anselin (1995). Akan tetapi, uji ketergantungan spasial hanya dapat dilakukan pada satu peubah. Oleh karena itu, data peubah ganda direduksi menjadi suatu peubah menggunakan analisis komponen utama (Johnson & Wichern, 2007). Pembobot spasial yang memenuhi uji indeks moran pada hipotesis H_1 (terdapat ketergantungan spasial antara lokasi) digunakan sebagai jarak ketergantungan spasial menggantikan jarak kedekatan geografis pada Chavent *et al.* (2018).

Oleh karena itu, pada penelitian ini akan dilakukan pengembangan analisis gerombol berhirarki menggunakan jarak peubah ganda dan jarak ketergantungan spasial pada data indikator makro sosial ekonomi di Kabupaten/Kota Provinsi Sulawesi Tengah. Adapun tujuan yang ingin dicapai dalam penelitian ini adalah menemukan gerombol Kabupaten/Kota Provinsi Sulawesi Tengah dengan mempertimbangkan kemiripan pola data indikator makro sosial ekonomi dan ketergantungan spasial di Kabupaten/Kota Provinsi Sulawesi Tengah tahun 2018.

2. Metodologi

2.1 Bahan dan Data

Data yang digunakan dalam penelitian ini (Tabel 1) adalah data indikator makro sosial ekonomi pada 13 Kabupaten/Kota Provinsi Sulawesi Tengah Tahun 2018 yang

berasal dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Tengah. Pemilihan peubah-peubah penelitian memperhatikan kelengkapan data dan kebutuhan analisis.

Tabel 1: Peubah yang digunakan pada penelitian.

Peubah	Keterangan	Satuan
X ₁	Jumlah kasus kriminalitas	/kasus
X ₂	Indeks pembangunan manusia (IPM)	%
X ₃	Pengeluaran rata-rata per kapita	Rp/kapita
X ₄	Kepadatan penduduk	%

(Sumber: BPS (2019))

2.2 Metode Penelitian

Adapun langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Menghitung matriks jarak peubah ganda (D_0)

Jarak *euclidean* digunakan untuk menghitung jarak peubah ganda (D_0) pada indikator makro sosial ekonomi sebagaimana ditunjukkan pada Tabel 1. Diberikan $x' = [x_1, x_2, \dots, x_p]$ dan $y' = [y_1, y_2, \dots, y_p]$ maka jarak *euclidian* adalah sebagai berikut (Johnson & Wichern, 2007).

$$d(x, y) = \sqrt{(x - y)'(x - y)}$$

2. Menghitung matriks jarak ketergantungan spasial (D_1)

Jarak ketergantungan spasial (D_1) dihitung melalui tiga tahapan:

a. Menghitung pembobot spasial

Pembobot spasial yang digunakan adalah *exponential distance*. Adapun persamaan pembobot *exponential distance* adalah sebagai berikut:

$$w_{ij} = \exp(-\alpha d_{ij})$$

dimana d_{ij} merupakan jarak *euclidean* menggunakan data *latitude* dan *longitude* Kabupaten/Kota Provinsi Sulawesi Tengah. Nilai α merupakan bilangan asli dan pada penelitian ini digunakan $\alpha = 6$ (Smith, 2014).

b. Melakukan analisis komponen utama

Indikator makro sosial ekonomi (Tabel 1) akan direduksi menjadi suatu peubah agar dapat dilakukan uji ketergantungan spasial melalui analisis komponen utama. Misalkan matriks ragam peragam (Σ) dari peubah X berukuran $n \times p$ dengan pasangan akar ciri dan vektor ciri $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ dimana $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ maka komponen utama ke- i memenuhi persamaan sebagai berikut.

$$W_i = e_i' X$$

Nilai dari proporsi kontribusi keragaman komponen utama ke- i , yaitu:

$$\text{Proporsi Komponen}(i) = \frac{\lambda_i}{\text{tr}(\Sigma)}$$

Semakin besar proporsi keragaman suatu komponen maka semakin besar keragaman dari peubah asal dapat dijelaskan (Johnson & Wichern, 2007).

c. Uji ketergantungan spasial

Uji ketergantungan spasial yang umumnya digunakan adalah indeks moran dengan prosedur pengujian sebagai berikut:

- Hipotesis
 $H_0: I = 0$; Tidak terdapat ketergantungan spasial antar lokasi
 $H_1: I \geq 0$; Terdapat ketergantungan spasial antar lokasi.
- Statistik Uji

dimana,

$$Z(I) = \frac{I - E(I)}{S_{E(I)}}$$

$$I = \frac{n}{\sum_i^n \sum_j^n w_{ij}} \frac{\sum_i^n \sum_j^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i^n (y_i - \bar{y})^2}$$

$$E(I) = -\frac{1}{n-1}$$

$$S_{E(I)} = \text{SQRT} \left[\frac{N^2 \sum_{ij} w_{ij}^2 + 3(\sum_{ij} w_{ij})^2 - N \sum_i (\sum_j w_{ij})^2}{(N^2 - 1)(\sum_{ij} w_{ij})^2} \right]$$

$$I = \frac{n}{\sum_i^n \sum_j^n w_{ij}} \frac{\sum_i^n \sum_j^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i^n (y_i - \bar{y})^2}$$

dengan

w_{ij} = bobot spasial antara lokasi ke-i dan ke-j

y_i / y_j = observasi ke-i/ke-j

\bar{y} = rata-rata observasi yang diamati

- Kesimpulan: Pengambilan keputusan H_0 ditolak atau ada ketergantungan antar lokasi jika $|Z_{\text{hitung}}| > Z_{\alpha/2}$ (Anselin, 1995).

Pembobot spasial yang memenuhi uji indeks moran pada hipotesis H_1 (terdapat ketergantungan spasial antara lokasi) digunakan sebagai jarak ketergantungan spasial (D_1).

3. Melakukan analisis gerombol berhirarki dengan ketergantungan spasial.

Analisis gerombol berhirarki dengan ketergantungan spasial sama dengan *Agglomerative Hierarchical Methods* menggunakan *wald linkage* hanya saja inputan jarak yang digunakan berbeda. Algoritma berhirarki dengan ketergantungan spasial adalah sebagai berikut:

- Hitunglah matriks jarak peubah ganda (D_0) dan jarak ketergantungan spasial (D_1). Inisiasi nilai $k = n$ sehingga terbentuk gerombol $P_k = \{C_1, C_2, \dots, C_k\}$. Pada tahapan ini diasumsikan *within-cluster inertia* ($\delta_\alpha(C_a, C_b) = 0$).
- Langkah berikutnya melakukan penggabungan gerombol awal untuk $k = n - 1, n - 2, n - 3, \dots, 2$ secara iteratif. Misalnya untuk $k=n-1$, pada seluruh kemungkinan kombinasi pasangan gerombol awal ($C_a \cup C_b$) dimana $C_a \subset P_n$ dan $C_b \subset P_n$ akan dihitung $\delta_\alpha(C_a, C_b)$. Kombinasi pasangan gerombol awal yang akan menjadi gerombol baru adalah pasangan gerombol yang memenuhi persamaan berikut:

$$\min_{C_a, C_b \in \mathcal{P}_{K+1}^\alpha} \delta_\alpha(C_a, C_b) = I_\alpha(C_a \cup C_b) - I_\alpha(C_a) - I_\alpha(C_b)$$

dimana

$$I_\alpha(C_a \cup C_b) = I_\alpha(C_k^\alpha) = (1 - \alpha) \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k} d_{0,ij}^2 + \alpha \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k} d_{1,ij}^2$$

dengan $\mu_k = \sum_{i \in C_k^\alpha} w_i$ dan w_i, w_j merupakan matrix bobot yang bernilai $1/n$. Setelah gerombol baru terbentuk proses dilanjutkan untuk $k=n-2, n-3, \dots, 2$.

- Berhenti ketika $K = 1$.

Nilai parameter pembobot jarak ketergantungan spasial (α) dapat diketahui dengan melakukan proses iteratif pada algoritma berhirarki dengan ketergantungan spasial sampai dengan diperoleh besaran *cut off pseudo inertia* antara jarak D_0 dan D_1 . Pada penelitian ini nilai α yang dicobakan adalah 0, 0.1, 0.2, ..., 1 (Chavent *et al.*, 2018).

4. Penentuan jumlah gerombol (k) terbaik.

Jumlah gerombol (k) ditentukan menggunakan jumlah *within sum of square* (wss) atau jumlah keragaman didalam gerombol. Diberikan gerombol C_k dengan nilai pusat μ_k maka jumlah wss adalah sebagai berikut:

$$\text{jumlah wss} = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

semakin kecil wss maka semakin baik gerombol dihasilkan (Kassambara, 2017).

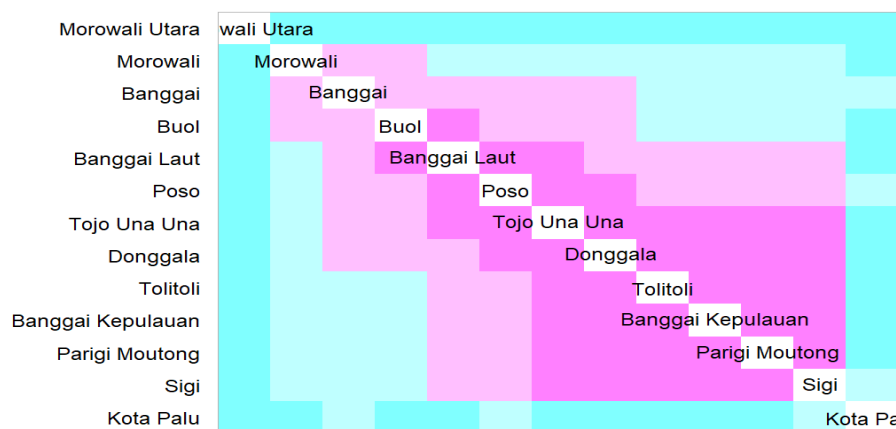
5. Interpretasi hasil penggerombolan.

Interpretasi hasil penggerombolan dilakukan menggunakan gambar peta Sulawesi Tengah dan data indikator makro sosial ekonomi.

3. Hasil dan Pembahasan

3.1 Menghitung Matriks Jarak Peubah Ganda (D_0)

Matriks jarak peubah ganda (D_0) dihitung menggunakan indikator makro sosial ekonomi yang terdiri dari X_1, X_2, X_3 dan X_4 sebagaimana ditunjukkan pada Tabel 1. Adapun hasil perhitungan jarak indikator makro sosial ekonomi adalah sebagai berikut.



Gambar 1: Jarak pada indikator makro sosial ekonomi

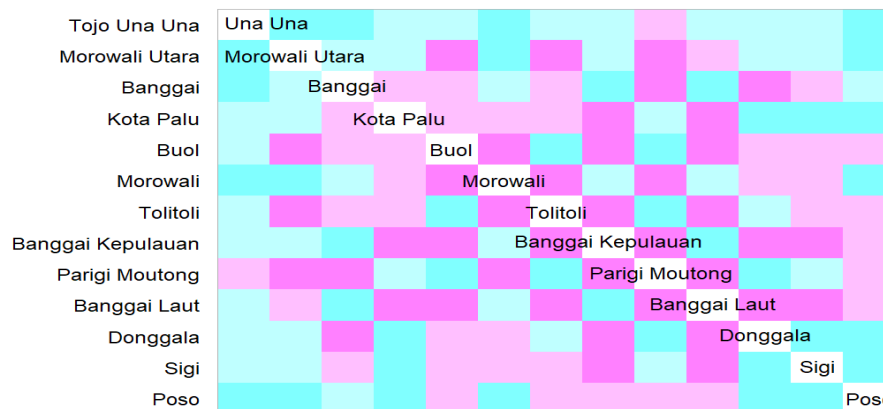
Jarak pada indikator makro sosial ekonomi diilustrasikan dengan perubahan degradasi warna. Semakin gelap maka semakin mirip pola indikator makro sosial ekonomi antar Kabupaten/Kota tersebut. Misalnya Kabupaten Morowali memiliki kesamaan pola indikator makro sosial ekonomi dengan Kabupaten Banggai dan Buol sebagaimana ditunjukkan pada Gambar 1.

3.2 Menghitung Matriks Jarak Ketergantungan Spasial (D_1)

Jarak ketergantungan spasial (D_1) dihitung melalui tiga tahapan yaitu pembentukan pembobot spasial, analisis komponen utama dan uji ketergantungan spasial.

a. Pembobot Spasial

Pembobot spasial *exponential distance* dihitung menggunakan *latitude* dan *longitude* Kabupaten/Kota Sulawesi Tengah dengan nilai $\alpha = 6$. Adapun hasil perhitungan pembobot spasial *exponential distance* adalah sebagai berikut.



Gambar 2: Pembobot spasial *exponential distance* dengan $\alpha = 6$.

Pembobot spasial *exponential distance* diilustrasikan juga dengan perubahan degradasi warna. Semakin terang maka semakin dekat Kabupaten/Kota tersebut. Misalnya Kota Palu memiliki bertetangga dengan Kabupaten Sigi, Parigi Moutong, Donggala dan Poso sebagaimana ditunjukkan pada Gambar 2.

b. Analisis Komponen Utama

Indikator makro sosial ekonomi direduksi menjadi suatu peubah yang merupakan kombinasi linier dari peubah tersebut sehingga dapat dilakukan uji ketergantungan spasial. Adapun hasil analisis komponen utama yang diperoleh adalah sebagai berikut:

Tabel 2: Proporsi Keragaman Komponen Utama

Components	PC1	PC2	PC3	PC4
Standard deviation	16570	1608	155	0,000000000001
Proportion of Variance	0,9906	0,00933	0,00009	0
Cumulative Proportion	0,9906	0,9999	1	1

Proporsi kumulatif keragaman menunjukkan seberapa besar suatu komponen dapat menjelaskan keragaman dari peubah asal. Proporsi kumulatif komponen

pertama (PC1) dapat menjelaskan 0,9906 atau 99,06% keragaman dari data sebagaimana ditunjukkan pada Tabel 2.

c. Uji Ketergantungan Spasial

Uji ketergantungan spasial dilakukan dengan menggunakan komponen pertama (PC1) hasil analisis komponen utama dan matriks pembobot spasial *exponential distance*. Adapun hasil uji indeks moran yang diperoleh adalah disajikan pada Tabel 3.

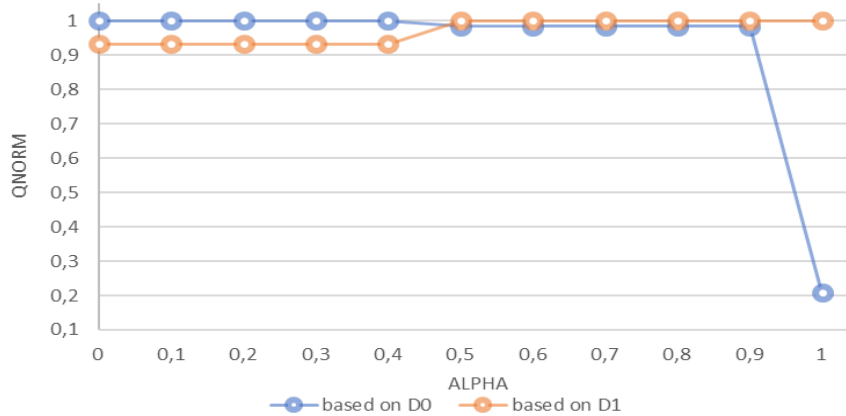
Tabel 3: Uji indeks moran

<i>Moran I Statistic</i>	<i>p-value</i>
0.3814328	0.0490509

Berdasarkan Tabel 3 diperoleh nilai *p-value* lebih kecil dari nilai 0.05 sehingga tolak hipotesis H_0 atau terdapat ketergantungan spasial pada data indikator makro sosial ekonomi antar Kabupaten/Kota Provinsi Sulawesi Tengah.

3.3 Analisis gerombol berhirarki dengan ketergantungan spasial

Analisis gerombol berhirarki dengan ketergantungan spasial dilakukan secara iteratif untuk menentukan pembobot jarak ketergantungan spasial (α) optimal untuk setiap jumlah gerombol (k). Misalkan untuk $k = 5$, diberikan jarak peubah ganda (D_0) dan jarak ketergantungan spasial (D_1) maka α yang optimal adalah sebagai berikut:



Gambar 3: Pembobot jarak ketergantungan spasil (α) pada $k = 5$.

Pembobot jarak ketergantungan spasial akan optimal pada $\alpha = 0.5$ sebagaimana ditunjukkan pada Gambar 3. Oleh karena itu, pengerombolan dengan berhirarki dengan jumlah gerombol $k = 5$ akan menggunakan $\alpha = 0.5$. Hal yang sama juga dilakukan untuk $k = 2,3,4$ dan 6 sehingga diperoleh nilai (α) terbaik berdasarkan jumlah gerombol seperti disajikan pada Tabel 4.

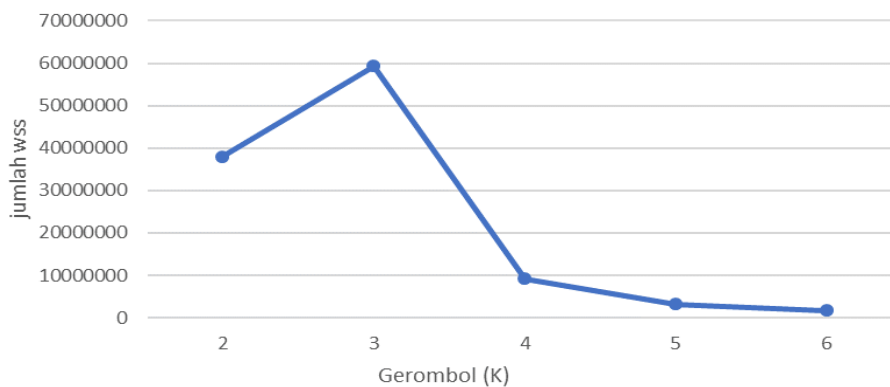
Tabel 4: Nilai α terbaik berdasarkan jumlah gerombol.

k	α
2	0.7
3	1.0
4	0.9
5	0.5
6	0.5

Simulasi jumlah gerombol terbaik akan menggunakan nilai α terbaik pada masing-masing jumlah gerombol sebagaimana ditunjukkan pada Tabel 4.

3.4 Penentuan jumlah gerombol (k) terbaik

Jumlah gerombol terbaik ditentukan berdasarkan keragaman didalam gerombol. Adapun hasil penentuan jumlah gerombol terbaik disajikan pada Gambar 4.

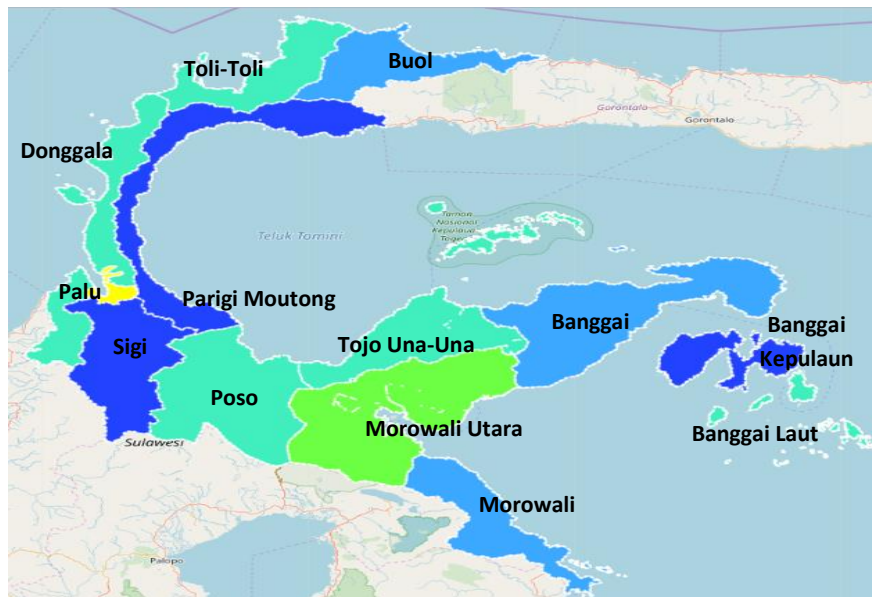


Gambar 4: Jumlah gerombol terbaik berdasarkan jumlah *within sum of square*.

Semakin kecil nilai *within sum of square* (wss) maka semakin homogen keragaman didalam gerombol. Jumlah gerombol 5 dengan $\alpha = 0.5$ merupakan jumlah gerombol optimal yang terbentuk karena perubahan jumlah wss telah cukup kecil pada jumlah k berikutnya sebagaimana ditunjukkan pada Gambar 4.

3.5 Interpretasi hasil penggerombolan

Analisis gerombol berhirarki dilakukan menggunakan jumlah gerombol dan nilai α terbaik yang telah diperoleh pada tahapan sebelumnya. Adapun hasil penggerombolan yang diperoleh disajikan pada Gambar 5. Gerombol Kabupaten/Kota yang ditunjukkan dengan kesamaan warna pada Gambar 5 dipengaruhi ketergantungan spasial dan tanpa ketergantungan spasial. Ketergantungan spasial terjadi pada Kabupaten/Kota yang berdekatan dan berada dalam gerombol yang sama. Misalnya Kabupaten Donggala, Toli-Toli, Tojo Una-Una, dan Poso sebagaimana ditunjukkan pada gambar tersebut. Sedangkan penggerombolan tanpa ketergantungan spasial terjadi pada Kabupaten/Kota yang berjauhan dan berada dalam gerombol yang sama. Hal ini terjadi karena pola penggerombolan berdasarkan indikator makro sosial ekonomi lebih dominan dari pada ketergantungan spasialnya. Misalnya Kabupaten Banggai, Buol, dan Morowali sebagaimana ditunjukkan pada Gambar 5.



Gambar 5: Gerombol berhirarki Kabupaten/Kota Provinsi Sulawesi Tengah menggunakan nilai jumlah gerombol dan nilai α terbaik.

Adapun interpretasi secara lengkap pada setiap gerombol disajikan pada Tabel 5.

Tabel 5: Interpretasi hasil penggerombolan Kabupaten/Kota di Sulawesi Tengah berdasarkan data indikator makro sosial ekonomi

Gerombol	Jumlah Kasus Kriminalitas	IPM	Pengeluaran Rata-Rata Per Kapita	Kepadatan Penduduk
Kabupaten Banggai Kepulauan, Parigi Moutong, dan Sigi	Tinggi	Rendah	Rendah	Tinggi
Kabupaten Banggai Laut, Tojo Una-Una, Poso, Donggala, dan Toli-Toli	Rendah	Rendah	Rendah	Tinggi
Kabupaten Buol, Banggai, dan Morowali	Tinggi	Tinggi	Tinggi	Rendah
Kabupaten Morowali Utara	Rendah	Tinggi	Tinggi	Rendah
Kota Palu	Tinggi	Tinggi	Tinggi	Tinggi

Penggerombolan Kabupaten/Kota di Sulawesi Tengah pada data indikator makro sosial ekonomi dipengaruhi oleh ketergantungan spasial sebagaimana ditunjukkan pada Tabel 5 dan Gambar 5.

3.6 Simpulan

Penggerombolan Kabupaten/Kota di Sulawesi Tengah pada data indikator makro sosial ekonomi dipengaruhi oleh ketergantungan spasial. Hal ini dapat dilihat pada Kabupaten/Kota yang berdekatan dan berada dalam gerombol yang sama. Analisis gerombol berhirarki dengan ketergantungan spasial pada Kabupaten/Kota Provinsi Sulawesi Tengah menghasilkan lima (5) gerombol antara lain 1) Kabupaten Banggai Kepulauan, Parigi Moutong, dan Sigi; 2) Kabupaten Banggai Laut, Tojo Una-Una, Poso, Donggala, dan Toli-Toli; 3) Kabupaten Buol, Banggai, dan Morowali; 4) Kabupaten Morowali Utara; dan 5) Kota Palu.

Ucapan Terima Kasih. Ucapan terima kasih disampaikan kepada Pimpinan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tadulako atas bantuan Pendanaan sehingga penelitian ini dapat terlaksana dengan baik.

Daftar Pustaka

- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2): 93–115.
- [BPS] Badan Pusat Statistik Provinsi Sulawesi Tengah. (2019). *Indikator Makro Ekonomi Provinsi Sulawesi Tengah Triwulan I Tahun 2019*. Palu (ID): Badan Pusat Statistik Provinsi Sulawesi Tengah.
- Chavent, M., Kuentz-Simonet, V., Labenne, A., & Saracco, J. (2018). ClustGeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4): 1799–1822.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (Vol. 6). New Jersey (US): Prentice hall Upper Saddle River, NJ.
- Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). STHDA.
- Smith, T. E. (2014). Spatial weight matrices. Retrieved November 1, 2019, from https://www.seas.upenn.edu/ese502/lab-content/extra_materials/SPATIAL%20WEIGHT%20MATRICES.pdf