

## Evaluasi Regresi Terklaster *Fuzzy* Spasial Simultan dengan Pendekatan Simulasi

Siti Hasanah<sup>1 \*</sup>, Muhammad Nur Aidi<sup>2</sup>, Anik Djuraidah<sup>3</sup>

Jl. Raya Dramaga, Kabupaten Bogor, 16680

<sup>1,2,3</sup>Statistika dan Sains Data, IPB University

\*e-mail: stathanah@apps.ipb.ac.id

*Diajukan: 18 Juni 2025, Diperbaiki: 13 Agustus 2025, Diterima: 6 Oktober 2025*

### Abstrak

Data spasial merupakan data yang memuat informasi yang berkaitan dengan karakteristik geografis suatu wilayah. Perkembangan data spasial yang mengarah pada data berskala besar membutuhkan metode analisis yang efisien dalam proses pengolahannya. Salah satu metode analisis yang dapat digunakan untuk mengolah data spasial berskala besar adalah *spatial fuzzy clustering*. Metode ini memungkinkan adanya penyesuaian bobot kelompok berdasarkan kemungkinan data, sehingga lebih mampu menangkap variasi lokal yang sebenarnya terjadi dalam data spasial. Metode *spatial fuzzy clustering* dengan penalti spasial, *Spatial Fuzzy Clustered Regression* (SFCR) dan tanpa penalti spasial, *Fuzzy Geographically Weighted Clustering Regression* (FGWCR) dievaluasi melalui simulasi pada penelitian ini. SFCR merupakan metode yang menggabungkan klusterisasi spasial dan pembentukan persamaan regresi secara simultan, sehingga waktu komputasi menjadi lebih efisien. FGWCR menghasilkan klaster yang mempertimbangkan kedekatan spasial dan kesamaan atribut sehingga efektif digunakan pada data spasial. Data dirancang sehingga terdapat 6 klaster dalam proses simulasi. Hasil simulasi menunjukkan metode SFCR lebih mampu mencerminkan keragaman data dan pembagian klaster dengan akurat. Nilai  $R^2$  untuk metode SFCR pada derajat *fuzziness* 2 dan autokorelasi spasial lemah, moderat, dan kuat berturut-turut yaitu 99.7%, 99.6%, dan 99.5%, sedangkan untuk metode FGWCR yaitu 98.5%, 98.6%, dan 98.1%. Keباikan model dievaluasi oleh nilai RMSE. Semakin kecil nilai RMSE maka model yang dihasilkan semakin baik. Nilai RMSE untuk metode SFCR pada derajat *fuzziness* 2 dan autokorelasi spasial lemah, moderat, dan kuat berturut-turut yaitu 0.30, 0.289, dan 0.298, sedangkan untuk metode FGWCR yaitu 0.659, 0.541, dan 0.551.

**Kata Kunci:** regresi terklaster simultan, klusterisasi spasial, *spatial fuzzy*, SFCR, FGWCR.

### Abstract

*Spatial data refers to data that contains information related to the geographical characteristics of a region. As spatial data evolves into large-scale datasets, efficient analytical methods are required for processing the data. One such method suitable for analyzing large-scale spatial data is spatial fuzzy clustering. This method allows for the adjustment of cluster weights based on data likelihood, making it more capable of capturing the actual local variations present in spatial data. In this study, two types of spatial fuzzy clustering methods were evaluated through simulation: the method with a spatial penalty, Spatial Fuzzy Clustered Regression (SFCR), and the method without a spatial penalty, Fuzzy Geographically Weighted Clustering Regression (FGWCR). SFCR is a method that combines spatial clustering and regression modeling simultaneously, resulting in more efficient computation time. FGWCR produces clusters by considering both spatial proximity and attribute similarity, making it effective for spatial data analysis. The data were designed to form six clusters during the simulation process. The simulation results showed that the SFCR method was more capable of accurately capturing data variation and cluster distribution. The  $R^2$  values for SFCR at a fuzziness degree of 2 and under weak, moderate, and strong spatial autocorrelation were 99.7%, 99.6%, and 99.5%, respectively, while the  $R^2$  values for FGWCR were 98.5%, 98.6%, and 98.1%. Model performance was evaluated using RMSE, where lower RMSE values indicate better performance. The RMSE values for the SFCR method at a fuzziness degree of*

2 and under weak, moderate, and strong spatial autocorrelation were 0.30, 0.289, and 0.298, respectively, while the RMSE values for the FGWCR method were 0.659, 0.541, and 0.551.

**Keywords:** Simultaneous clustered regression, spatial clustering, spatial fuzzy, SFCR, FGWCR.

## 1 Pendahuluan

Perkembangan metode analisis untuk data spasial berskala besar salah satunya dengan metode *spatial fuzzy clustering*. Metode ini memungkinkan adanya penyesuaian bobot kelompok berdasarkan kemungkinan data, sehingga lebih mampu menangkap variasi lokal yang sebenarnya terjadi dalam data spasial. Data spasial merupakan data yang memuat informasi atribut dan lokasi geografis dengan sistem koordinat tertentu sebagai dasar referensinya [1]. Menurut Cressie dalam Djuraidah [1], data geospasial terdiri dari tiga jenis, yaitu pola titik, data geostatistik, dan data area. Pemodelan data area perlu memperhatikan efek spasial seperti dependensi spasial dan keragaman spasial. Daerah yang berdekatan diharapkan lebih mirip dibandingkan daerah yang berjauhan, merupakan gagasan dasar dari dependensi spasial. Keragaman spasial terdiri atas dua macam yaitu ketidakstabilan struktural dan heteroskedastisitas [1].

Ketidakstabilan struktural yaitu parameter struktural/koeffisien regresi yang tidak stabil pada semua lokasi menyebabkan koefisien regresi yang berbeda-beda pada setiap lokasi amatan. Pendekatan persamaan regresi spasial untuk mengatasi hal ini adalah *Geographically Weighted Regression* (GWR). Seluruh amatan pada persamaan ini digunakan dalam pendugaan dengan bobot yang berbeda sehingga menghasilkan persamaan regresi lokal untuk setiap lokasi. Pada persamaan GWR yang areanya luas (data lokasi banyak) akan lebih mudah diterapkan bila jarak terbatas sehingga sering digunakan *bandwith*. Namun demikian persamaan GWR memiliki beberapa kekurangan yaitu cenderung menghasilkan koefisien ekstrim pada dataset dengan kerapatan spasial rendah dan data area lebih sensitif terhadap koefisien ekstrim tersebut [2]. Pendekatan GWR dengan regularisasi yaitu metode pengurangan parameter dan penyederhanaan persamaan seperti Lasso dan Ridge telah dikembangkan untuk mengakomodir kekurangan tersebut namun pendekatan tersebut tetap tidak dapat mencapai efisiensi secara komputasi [3]. Bila dataset yang digunakan berukuran besar maka persamaan GWR yang terbentuk akan sangat banyak dan kompleks. Terdapat alternatif lain untuk mengatasi hal tersebut yakni dilakukan pengelompokan yang terdiri atas dua macam metode. Pertama pengelompokan tanpa penalti yaitu *Fuzzy Geographically Weighted Clustering Regression* (FGWCR), dan kedua pengelompokan dengan menyertakan fungsi penalti pada *Spatial Fuzzy Clustered Regression* (SFCR).

Modifikasi persamaan GWR dengan penggerombolan *fuzzy* merupakan salah satu alternatif untuk mengatasi kompleksitas persamaan pada dataset berukuran besar. FGWCR menerapkan

efek geografis pada algoritma penggerombolan *fuzzy*, sehingga memiliki keunggulan dalam mengatasi pencilan [4]. Penelitian oleh [5] menunjukkan model berbasis ketergantungan spasial lebih andal dalam stabilitas estimasi. FGWC sering dimanfaatkan dalam analisis geodemografi karena dapat mengakomodasi pengaruh kedekatan geografis antar entitas spasial sekaligus menyempurnakan kinerja algoritma *fuzzy c-means* standar[6][7].

Sugasawa dan Murakami [3] memperkenalkan metode baru yang menggabungkan persamaan regresi spasial dengan penggerombolan *fuzzy* yaitu SFCR. Perbedaan metode ini dari metode-metode sebelumnya yaitu proses pembentukan persamaan regresi dan penggerombolan dilakukan secara simultan. Hal tersebut bertujuan untuk menghasilkan penggerombolan spasial yang sesuai dengan struktur regresinya dan proses komputasi menjadi lebih efisien. Suatu amatan dengan lokasi yang bertetangga memungkinkan untuk memiliki karakteristik yang mirip sehingga dapat menjadi satu gerombol. Fungsi penalti yang didasari oleh persamaan Potts digunakan untuk mendorong struktur tersebut. Penentuan gerombol pada persamaan SFCR ditentukan oleh suatu kriteria informasi. Melalui Teknik penggerombolan tersebut hasil pendugaan menjadi lebih stabil dan lebih mudah diinterpretasikan dibandingkan dengan metode GWR.

Pembobotan berdasarkan persamaan karakteristik antarwilayah dengan menerapkan metode *fuzzy* terklaster telah diteliti oleh Andrytiarandy pada tahun 2017. Hasil penelitiannya menunjukkan metode GWR dengan penggerombolan *fuzzy* lebih unggul dibandingkan dengan penggunaan persamaan GWR saja[8]. Namun pada penelitian tersebut belum menggunakan metode *fuzzy* dengan efek geografis/spasial. Pendekatan klaster dalam distribusi spasial juga diteliti oleh [9] tanpa pembentukan model regresi dan oleh [10] dengan pembentukan regresi nonsimultan. Penelitian ini merupakan kajian yang baru karena belum ada penelitian yang serupa sebelumnya. Oleh karena itu penelitian ini bertujuan untuk mengevaluasi metode *spatial fuzzy clustering* yaitu SFCR dan FGWCR melalui simulasi.

## 2 Metode Penelitian

### 2.1 Data

Data yang digunakan dalam penelitian ini merupakan data yang dibangkitkan melalui proses simulasi dengan 539 titik lokasi kecamatan dalam Provinsi Jawa Barat. Domain lokasi yaitu  $\{s_i = (u_i, v_i) | u_i \in [-7.7583, -5.99626], v_i \in [106.44792, 108.79253]\}, i = 1, \dots, 539$ . Peubah penjelas berjumlah 8 peubah yang dibangkitkan melalui proses Gaussian spasial dengan eksponensial isotropik [11].  $z_k(s_i)$  merupakan nilai dari proses Gaussian spasial ke-  $k$  di lokasi  $(s_i)$  dengan

mean nol dan struktur kovarian spasial isotropik  $\exp(-\frac{\|s_i - s_j\|}{\eta})$ ,  $\eta$  mengontrol autokorelasi spasial lemah ( $\eta = 0,2$ ), moderat ( $\eta = 0,6$ ), dan kuat ( $\eta = 1$ ),  $k = 1, \dots, 8$ .

## 2.2 SFCR

Sugasawa dan Murakami [3] memperkenalkan metode baru untuk menduga pola hubungan antar peubah yang tergerombol secara spasial. Kombinasi penggerombolan dengan fungsi penalti yang terinspirasi dari persamaan Potts mendorong terbentuknya gerombol pada lokasi yang bertetangga. Ide yang mendasari penggabungan regresi spasial dengan penggerombolan adalah lokasi geografis yang dapat dibagi kedalam beberapa gerombol, dimana lokasi bertetangga dengan karakteristik yang sama dapat menjadi satu gerombol. Teknik penggerombolan ini menyebabkan hasil pendugaan menjadi lebih stabil secara numerik dan lebih mudah diinterpretasikan dibandingkan dengan persamaan GWR [3]. SFCR menggunakan algoritma iteratif dan dapat diimplementasikan pada data berukuran besar, serta merupakan pendekatan efektif terhadap regresi spasial dengan kondisi koefisien yang tidak stasioner.

Keanggotaan di lokasi yang berdekatan cenderung memiliki keanggotaan yang sama didorong oleh fungsi penalti yang didasarkan pada proses spasial yang dikenal sebagai persamaan Potts. Fungsi probabilitas gabungan dari persamaan Potts diberikan oleh:

$$\pi(g_1, \dots, g_n | \phi) \propto \exp(\phi \sum_{i < j} w_{ij} I(g_i = g_j)) \quad (1)$$

dengan  $w_{ij}$  adalah fungsi bobot,  $\phi$  mengontrol kekuatan autokorelasi spasial, dan  $g_1, \dots, g_n$  adalah parameter keanggotaan. Suatu fungsi objektif yang akan dimaksimumkan secara iteratif merupakan fungsi *penalized likelihood* sebagai berikut:

$$Q(\theta, g) = \sum_{i=1}^n \log f(y_i | x_i; \theta_{gi}) + \phi \sum_{i < j} w_{ij} I(g_i = g_j) \quad (2)$$

dengan  $\theta$  adalah parameter struktural yang mencerminkan heterogenitas spasial. Proses iterasi akan memperbaharui parameter keanggotaan dan parameter lainnya.

Penggerombolan yang bersifat *hard clustering* kemungkinan dapat membatasi akurasi estimasi. Selanjutnya untuk mengatasi kesulitan ini, [3] menggunakan pengelompokan *fuzzy*, yang memungkinkan ketidakpastian pengelompokan dengan memperkenalkan bobot yang disempurnakan serta ditentukan oleh fungsi *likelihood*. Selain itu, dipertimbangkan pula probabilitas bahwa lokasi ke-  $i$  termasuk dalam grup  $g$  dengan:

$$\pi_{ig} = \frac{[f(y_i | x_i; \theta_g) \exp\{\phi \sum_{j=1; j \neq i}^n w_{ij} I(g = g_j)\}]^\delta}{\sum_{g'=1}^G [f(y_i | x_i; \theta_{g'}) \exp\{\phi \sum_{j=1; j \neq i}^n w_{ij} I(g' = g_j)\}]^\delta} \quad (3)$$

dengan  $\delta$  mengontrol derajat *fuzziness*. Dua tuning parameter yaitu jumlah gerombol ( $G$ ) dan  $\phi$  yang mengontrol kekuatan dependensi spasial. Jumlah gerombol ditentukan berdasarkan informasi prior dari data dengan suatu kriteria informasi:

$$IC(G) = -2 \sum_{i=1}^n \log f(y_i | x_i; \widehat{\theta}_{gl}) + c_n \dim(\theta) \quad (4)$$

dengan  $c_n$  adalah konstanta yang tergantung pada ukuran contoh  $n$  dan  $\dim(\theta)$  menunjukkan dimensi dari  $\theta$  yang bergantung pada  $G$ . Secara khusus, digunakan  $c_n = \log n$  yang mengarah pada kriteria BIC. Nilai  $G$  yang sesuai dipilih sebagai  $\hat{G} = \operatorname{argmin}_{G \in \{G_1, \dots, G_L\}} IC(G)$ , di mana  $G_1, \dots, G_L$  adalah kandidat dari  $G$ .

### 2.3 FGWCR

FGWC merupakan modifikasi algoritma FCM dengan menambahkan langkah tambahan pada setiap iterasi, yaitu menerapkan pembobotan pada nilai keanggotaan gerombol. Pembobotan tersebut merupakan pengaruh geografis yang akan mempengaruhi nilai pusat kluster [12]. Berbeda dengan metode regresi klusterisasi *fuzzy* oleh [13] yang hanya menggunakan *fuzzy loading* tanpa memasukkan pengaruh geografis. Metode ini mengimplementasikan efek penurunan jarak melalui faktor pembobotan untuk menganalisis penggerombolan data geo-demografi. Setiap area diberikan nilai keanggotaan pada setiap gerombol sehingga membantu mengatasi masalah kekeliruan ekologi [14]. Metode ini mengadopsi prinsip interaksi spasial dengan menggabungkannya ke dalam pembobotan keanggotaan, menyesuaikan nilai keanggotaan gerombol berdasarkan efek ketetanggaan sehingga menghasilkan gerombol yang “sadar geografis” [14]. Fungsi objektif dapat dituliskan sebagai berikut:

$$J_{FGWC}(U, V; X) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^\delta |v_i - x_k|^2 \quad (5)$$

dengan  $\delta$  adalah derajat *fuzziness*,  $v_i$  pusat gerombol,  $u_{ik}$  adalah derajat keanggotaan dari objek terhadap pusat gerombol,  $n$  adalah jumlah objek,  $c$  adalah jumlah gerombol,  $x_k$  merupakan titik data. Secara umum derajat *fuzziness* yang optimal digunakan bernilai 2, yaitu saat terjadi keseimbangan antara *separability* (kejelasan kluster) dengan *fuzziness* [15]. Derajat *fuzziness* yang mendekati nilai 1 berefek menjadi *hard clustering*, sedangkan bila terlalu besar keanggotaan menjadi kabur sehingga tidak ada perbedaan antar kluster [16]. Efek ketetanggaan yang digabungkan kedalam algoritma juga akan mempengaruhi pusat gerombol. Pusat gerombol dapat ditentukan dengan:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^\delta x_k}{\sum_{k=1}^n u_{ik}^\delta} \quad (6)$$

Penggabungan efek ketetanggaan tersebut menyebabkan keanggotaan gerombol terhadap suatu area tertentu terus diperbaharui secara iteratif. Berikut formula keanggotaan gerombol:

$$\mu'_i = \alpha\mu_i + \beta \frac{1}{A} \sum_j^n w_{ij} \mu_j \quad (7)$$

dengan  $\mu'_i$  adalah keanggotaan gerombol baru untuk area-  $i$ ,  $\mu_i$  merupakan keanggotaan yang lama, nilai  $\alpha + \beta = 1$  yang merupakan peubah untuk mempengaruhi proporsi antara keanggotaan awal dengan keanggotaan terboboti, serta  $w_{ij}$  adalah bobot spasial. Bobot keanggotaan selanjutnya menjadi dasar untuk pembobotan persamaan regresi FGWCR.

## 2.4 Prosedur Simulasi

Prosedur simulasi dilakukan dengan memodifikasi penelitian Sugasawa dan Murakami (2021) melalui langkah-langkah sebagai berikut:

1. Menentukan batas 6 klaster  $T_{jk}$  dan membangkitkan koefisien regresi sebagai berikut:

$$T_{jk} = \{s_i = (u_i, v_i) | p_{1j} < v_i \leq p_{1,j+1}, p_{2k} < u_i \leq p_{2,k+1}\} \quad i = 1, \dots, 539; \quad j = 0, 1, 2; \quad k = 0, 1$$

$$p_{1j} = \min(s_1) - 0,001 + j \frac{(\max(s_1) - \min(s_1) + 0,002)}{3}, \quad j = 0, 1, 2, 3$$

$$p_{2k} = \min(s_2) - 0,001 + k \frac{(\max(s_2) - \min(s_2) + 0,002)}{2}, \quad k = 0, 1, 2$$

$$\beta_0(s_i) = 0,2(p_{1j} + p_{2k})$$

$$\beta_1(s_i) = 0,01(p_{1j}^2 + p_{2k}^2)$$

$$\beta_2(s_i) = -0,01 + 0,02(p_{1j} + p_{2k})$$

$$\beta_3(s_i) = |0,1p_{1j} + 0,2p_{2k}|$$

$$\beta_4(s_i) = -0,01p_{1j} \cdot p_{2k}$$

$$\beta_5(s_i) = -0,0001p_{2k} \cdot p_{1j}^2$$

$$\beta_6(s_i) = |\sin(p_{1j} \cdot p_{2k})|$$

$$\beta_7(s_i) = |\cos(p_{1j} + p_{2k})|$$

$$\beta_8(s_i) = 0,0001(-p_{2k}^2 + p_{1j}^2)$$

$$\sigma(s_i) = 0,005 + 0,002|-p_{1j} + p_{2k}|$$

$$\varepsilon(s_i) \sim N(0,1)$$

2. Membangkitkan 8 peubah penjelas dengan ketentuan sebagai berikut,

$$x_t = \Phi z_t, \quad t = 1, \dots, 5, 7$$

$$x_6 = \Phi(0,75z_2 + \sqrt{1 - 0,75^2}z_6)$$

$$x_8 = \Phi(0,9z_3 + \sqrt{1 - 0,9^2}z_8)$$

$$\text{Cov}(z_k(s_i), z_k(s_j)) = \exp\left(-\frac{\|s_i - s_j\|}{\eta}\right), \quad k = 1, \dots, 8; \quad i \neq j; \quad i, j = 1, \dots, 539$$

dengan  $\Phi$  merupakan fungsi distribusi kumulatif normal standar.

3. Menentukan nilai peubah respon dari persamaan regresi

$$y_i = \beta_0(s_i) + \beta_1(s_i)x_1 + \dots + \beta_8(s_i)x_8 + \sigma(s_i)\varepsilon(s_i), \quad i = 1, 2, \dots, n \quad (8)$$

4. Mendeteksi efek keragaman spasial. Pengujian dapat dilakukan dengan menggunakan uji Breusch-Pagan (Anselin 1988) dengan hipotesis nol ragam homogen dan statistik uji sebagai berikut:

$$BP = \frac{1}{2} \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f} \quad (9)$$

dengan  $f_i = \left( \frac{\varepsilon_i^2}{\sigma^2} - 1 \right)$ ,  $\mathbf{Z}$  adalah vektor amatan peubah dependen,  $\varepsilon_i$  kuadrat tengah galat untuk pengamatan ke-  $i$ , dan  $\sigma$  merupakan ragam dari  $\varepsilon_i$ . Pengambilan keputusan pada uji Breusch-Pagan yaitu apabila nilai  $BP > \chi^2_{(\alpha; p-1)}$  maka  $H_0$  ditolak.

5. Menghitung matriks pembobot spasial dengan cara menghitung jarak antar lokasi melalui formula jarak Euclidean dan bobot  $k$ -tetangga terdekat ( $k = 5$ ) yaitu:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (10)$$

dengan  $u_i$  merupakan kordinat lintang lokasi ke-  $i$ ,  $u_j$  adalah kordinat lintang lokasi ke-  $j$ ,  $v_i$  merupakan kordinat bujur lokasi ke-  $i$ , dan  $v_j$  adalah kordinat bujur lokasi ke-  $j$ . Bobot  $k$ -tetangga terdekat ditentukan dengan mengurutkan jarak  $d_{ij}$ ,  $d_{ij(1)} \leq d_{ij(2)} \leq \dots \leq d_{ij(n-1)}$ . Kemudian ditentukan  $k$ -lokasi tetangga terdekat dari lokasi-  $i$  yaitu  $N_k(i) = \{j(1), j(2), \dots, j(k)\}$  dengan nilai  $k=1, \dots, n-1$ .

6. Menerapkan prosedur SFCR dengan menentukan kombinasi parameter inisiasi  $\delta$  dan  $g^0$ . Setiap kombinasi yang konvergen dilakukan 15 kali hingga didapatkan 180 persamaan regresi.

a) Menghitung bobot keanggotaan gerombol dengan Persamaan (3)

b) Melakukan *update* parameter  $\theta_{(k)}$  dan  $g_{(k)}$  untuk  $g = 1, \dots, G$  melalui formula:

$$\theta_g^{k+1} = \operatorname{argmax}_{\theta_g} \sum_{i=1}^n \pi_{ig}^{(k)} \log f(y_i | x_i; \theta_g) \quad (11)$$

$$g_i^{k+1} = \operatorname{argmax}_{g \in \{1, \dots, G\}} \pi_{ig}^{(k)} \quad (12)$$

c) Ulangi dari langkah 6a hingga tercapai nilai konvergen dengan toleransi  $\varepsilon = 10^{-6}$  yaitu:

$$\Delta^{k+1} = \frac{|Q_k(\theta, g) - Q_{k+1}(\theta, g)|}{Q_{k+1}(\theta, g)} \quad (13)$$

$\Delta^{k+1}$  adalah selisih iterasi ke-  $k+1$  dan iterasi ke-  $k$ ,  $Q_k(\theta, g)$  merupakan fungsi objektif penalti *likelihood* ke-  $k$ , dan  $Q_{k+1}(\theta, g)$  adalah fungsi objektif penalti *likelihood* ke-  $k+1$

7. Menerapkan prosedur FGWCR dengan menentukan kombinasi parameter inisiasi  $\delta$  dan jumlah klaster. Setiap kombinasi yang konvergen dilakukan 15 kali hingga didapatkan 180 persamaan regresi
  - a) Menghitung pusat klaster dengan Persamaan (6) untuk memberikan posisi representatif dari tiap klaster dalam ruang data, dengan mempertimbangkan derajat keanggotaan *fuzzy*
  - b) Melakukan *update* matriks keanggotaan menggunakan informasi spasial dengan Persamaan (7) lalu menyimpan nilai fungsi objektif dan pembentukan regresi
  - c) Ulangi dari langkah 7a hingga tercapai konvergensi yaitu  $|u_{ik} - u_{ik-1}| < \varepsilon$ , dengan  $\varepsilon = 10^{-6}$
8. Mengevaluasi metode SFCR dan FGWCR dengan menghitung nilai MSE dan MAE dugaan koefisien regresi serta RMSE dan  $R^2$  dari dugaan peubah respon,

$$MSE = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n} \quad (14)$$

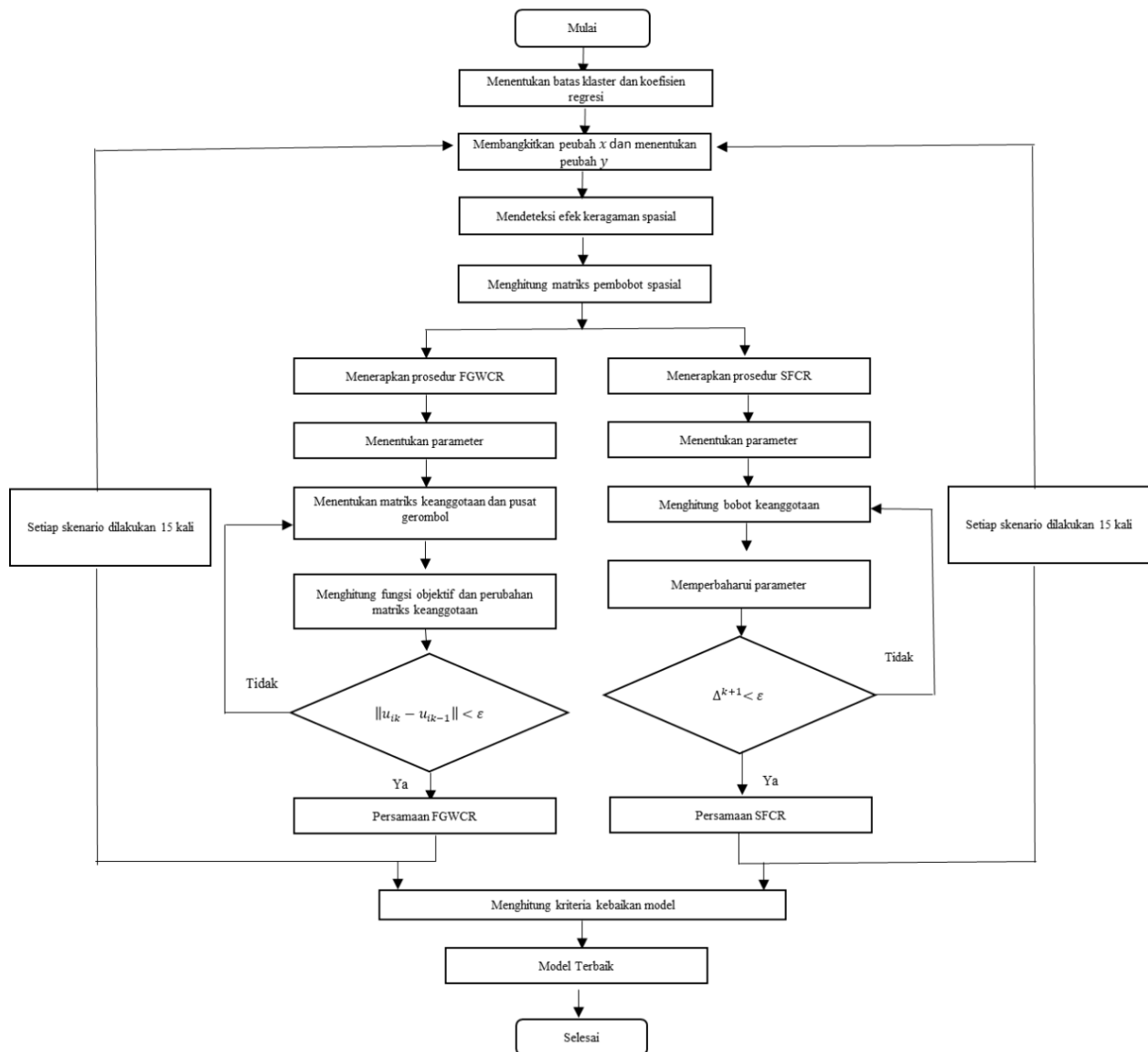
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$RMSE = \sum_{i=1}^n \sqrt{\frac{(y_i - \hat{y}_i)^2}{n}} \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad [17] \quad (17)$$

dengan  $y_i$  adalah nilai aktual ke-  $i$ ,  $\hat{y}_i$  nilai prediksi ke-  $i$ ,  $\bar{y}_i$  merupakan rata-rata nilai aktual, dan  $n$  merupakan jumlah data.





**Gambar 1.** Diagram alir prosedur simulasi

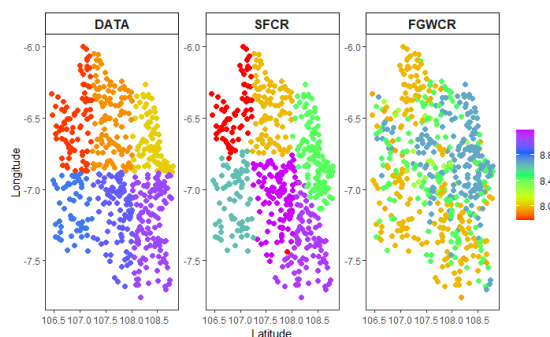
### 3 Hasil dan Pembahasan

Hasil deteksi efek keragaman spasial terhadap 539 data bangkitan seluruh skenario menunjukkan bahwa  $H_0$  ditolak dan ragam tidak homogen. Ketidakhomogenan ragam menunjukkan ketidakstabilan struktural yang menyebabkan koefisien regresi berbeda beda pada setiap lokasi amatan / klaster. Hal ini diatasi dengan metode klaster spasial melalui simulasi. Simulasi terhadap 6 klaster dengan nilai-nilai inisiasi parameter dan hasil uji efek keragaman spasial ditunjukkan oleh Tabel 1.

Tabel 1. Skenario kombinasi inisiasi parameter dan hasil uji efek keragaman spasial

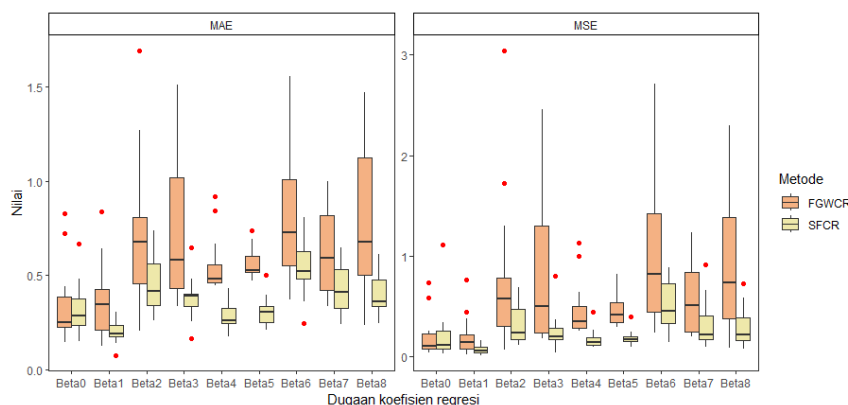
| Skenario | $\eta$ | $\delta$ | $p$ -value             | Kesimpulan                     |
|----------|--------|----------|------------------------|--------------------------------|
| Lemah    | 0,2    | 1,1      | $5,53 \times 10^{-26}$ | Terdapat heterogenitas spasial |
|          |        | 1,5      | $3,78 \times 10^{-4}$  |                                |
|          |        | 2        | $3,78 \times 10^{-4}$  |                                |
|          |        | 2,5      | $1,52 \times 10^{-11}$ |                                |
| Moderat  | 0,6    | 1,1      | $3,83 \times 10^{-16}$ | Terdapat heterogenitas spasial |
|          |        | 1,5      | $3,03 \times 10^{-6}$  |                                |
|          |        | 2        | $5,67 \times 10^{-6}$  |                                |
|          |        | 2,5      | $6,93 \times 10^{-7}$  |                                |
| Kuat     | 1      | 1,1      | $4,00 \times 10^{-13}$ | Terdapat heterogenitas spasial |
|          |        | 1,5      | $1,88 \times 10^{-5}$  |                                |
|          |        | 2        | $4,00 \times 10^{-13}$ |                                |
|          |        | 2,5      | $6,93 \times 10^{-2}$  |                                |

Kombinasi antara autokorelasi spasial dan derajat *fuzziness* menghasilkan 360 persamaan regresi. Autokorelasi spasial lemah (0,2) dan derajat *fuzziness* 1,1 yang diterapkan pada data simulasi menunjukkan hasil yang berbeda pada metode SFCR dan FGWCR. Visualisasi salah satu koefisien regresi (Beta 5) data bangkitan menunjukkan terdapat 6 kluster dan metode SFCR dapat menangkap keragaman tersebut pada dugaan koefisien regresinya. Terdapat 6 kluster yang dihasilkan oleh metode SFCR dan 3 kluster pada metode FGWCR seperti yang ditunjukkan oleh Gambar 2. Metode SFCR mampu mempertahankan struktur spasial yang cukup mirip dengan data bangkitan. Warna-warni kluster masih menunjukkan konsistensi spasial seperti, kluster berwarna merah dan oranye tetap dominan di utara serta warna ungu di selatan. Hal ini menunjukkan bahwa pendekatan spasial dengan penalti dalam SFCR cukup efektif menjaga struktur lokal, sekaligus menghasilkan segmentasi yang teratur dan jelas. Koefisien regresi pada metode FGWCR juga terlihat bergerombol namun tidak tersegmentasi dengan jelas seperti data bangkitan. Jumlah kluster yang dapat diidentifikasi metode ini pun jauh lebih sedikit. Hal ini dikarenakan FGWCR tidak menerapkan fungsi penalti dalam metodenya sehingga jumlah kluster yang terbentuk hanya berdasarkan kekuatan struktur lokal yang dominan.



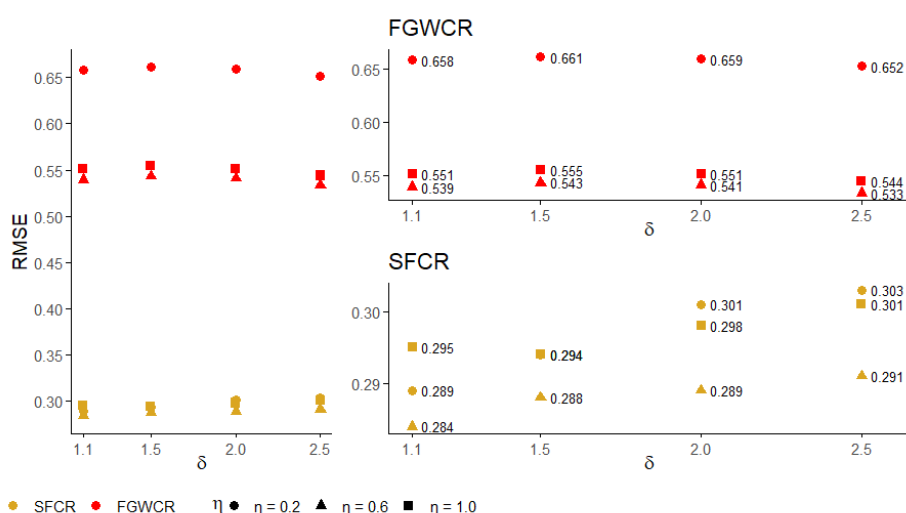
**Gambar 2.** Sebaran dugaan koefisien regresi (beta 5) dengan  $\eta = 0,2$  dan derajat *fuzziness* 1,1. Seluruh dugaan koefisien regresi kedua metode dievaluasi oleh nilai MSE dan MAE. Semakin kecil nilai keduanya, semakin dekat dugaan parameter dengan nilai sebenarnya dan model semakin baik [18]. Nilai MSE dan MAE tersebut divisualisasikan oleh diagram kotak garis pada Gambar

3. Secara umum metode SFCR menunjukkan nilai MSE dan MAE yang lebih kecil dan stabil hampir pada keseluruhan skenario. Hal ini menunjukkan metode SFCR lebih konsisten dan presisi dalam memperkirakan parameter.



**Gambar 3.** Diagram kotak garis dugaan koefisien regresi ( $\eta = 0,2$  dan derajat *fuzziness* 1,1) Adanya pencilan pada metode FGWCR terutama pada beta 0 dan beta 4 serta nilai MSE yang besar pada beta 3 dan beta 6 menunjukkan metode FGWCR kurang stabil dalam menduga parameter ini.

Persamaan regresi yang dihasilkan kedua metode dievaluasi oleh nilai koefisien determinasi ( $R^2$ ) dan RMSE dari dugaan peubah respon. Nilai  $R^2$  yang semakin mendekati 1 atau 100% menunjukkan model yang dihasilkan semakin baik dalam menjelaskan keragaman data. Sebaliknya untuk nilai RMSE, semakin kecil nilainya maka semakin baik dan prediksi model semakin akurat.



**Gambar 4.** Diagram pencar RMSE

Skenario lemah dengan derajat *fuzziness* 1,1 menghasilkan nilai RMSE 0,289 untuk metode SFCR dan 0,658 pada metode FGWCR. Keseluruhan nilai RMSE dapat dilihat pada Gambar 4. Nilai RMSE metode SFCR konsisten lebih kecil dibandingkan metode FGWCR. Metode SFCR

mengabungkan *fuzzy logic*, model regresi berbasis klaster, dan pengaruh spasial sehingga model yang dihasilkan lebih fleksibel dan adaptif terhadap pola data, terutama jika terdapat struktur lokal. Penyesuaian variasi lokal dengan klaster spasial dapat menekan galat yang berimplikasi semakin menurunnya nilai RMSE. Terdapat pola nilai RMSE yang menurun seiring meningkatnya derajat *fuzziness* pada metode FGWCR. Kondisi tersebut berkebalikan dengan metode SFCR, semakin tinggi nilai  $\delta$  semakin tinggi pula nilai RMSE. Pada metode FGWCR peningkatan nilai  $\delta$  menyebabkan keanggotaan  $u_{ik}$  lebih tersebar dan data amatan dapat diberi bobot pada beberapa klaster sehingga pengaruh pencilaan cenderung berkurang. Hal ini dapat menyebabkan model lebih kekar, kesesuaian model lebih baik, dan RMSE menurun.  $\delta$  meningkatkan pengaruh fungsi likelihood dan spasial secara eksponensial pada metode SFCR. Saat  $\delta$  meningkat keanggotaan menjadi lebih tegas mendekati *hard clustering* sehingga dapat menyebabkan nilai RMSE meningkat saat terjadi ketidakcocokan keanggotaan.

Hasil simulasi menunjukkan bahwa nilai RMSE tertinggi terjadi pada tingkat autokorelasi spasial rendah ( $\eta = 0,2$ ). Hal ini mengindikasikan lemahnya pemanfaatan informasi spasial ketika kekuatan spasial rendah. Sebaliknya pada  $\eta = 0,6$  model mencapai performa terbaik, menunjukkan adanya keseimbangan optimal antara ketelitian lokal dan kontribusi spasial. Kinerja kembali menurun pada  $\eta = 1$  akibat dominasi pengaruh spasial yang cenderung melunakkan respon model terhadap variasi lokal.

Nilai  $R^2$  pada metode SFCR secara konsisten lebih tinggi dibandingkan FGWCR pada seluruh kombinasi parameter, menunjukkan kemampuan model yang lebih baik dalam menjelaskan variabilitas data. Tabel 2 menunjukkan bahwa pada SFCR, peningkatan parameter  $\eta$  justru menurunkan  $R^2$ , mengindikasikan bahwa kekuatan spasial yang terlalu tinggi dapat mengurangi fleksibilitas model.

Tabel 2.  $R^2$  persamaan regresi

|          |     | $\eta$ |  | 0,2     |         | 0,6     |         | 1       |         |
|----------|-----|--------|--|---------|---------|---------|---------|---------|---------|
|          |     |        |  | SFCR    | FGWCR   | SFCR    | FGWCR   | SFCR    | FGWCR   |
| $\delta$ | 1,1 |        |  | 0,99720 | 0,98548 | 0,99628 | 0,98625 | 0,99479 | 0,98063 |
|          | 1,5 |        |  | 0,99709 | 0,98533 | 0,99619 | 0,98610 | 0,99487 | 0,98038 |
|          | 2   |        |  | 0,99696 | 0,98544 | 0,99617 | 0,98624 | 0,99468 | 0,98062 |
|          | 2,5 |        |  | 0,99692 | 0,98572 | 0,99611 | 0,98662 | 0,99456 | 0,98114 |

Sementara pada FGWCR, nilai  $R^2$  tertinggi dicapai pada  $\eta = 0,6$ , menunjukkan bahwa kekuatan spasial yang moderat memberikan keseimbangan antara regularisasi dan akurasi lokal. Kedua metode dalam persamaan regresinya mampu menjelaskan keragaman respon melalui peubah penjelasnya dengan sangat baik.

## 4 Simpulan

1. Secara konsisten metode SFCR memiliki nilai  $R^2$  yang lebih tinggi dibandingkan metode FGWCR pada seluruh kombinasi parameter  $\eta$  dan  $\delta$ , menunjukkan bahwa SFCR memiliki kemampuan yang lebih baik dalam menjelaskan variasi data.
2. Nilai RMSE metode SFCR konsisten berada di bawah metode FGWCR menunjukkan metode SFCR lebih akurat dalam memprediksi model.
3. Pengaruh peningkatan  $\eta$  dan  $\delta$  terhadap data simulasi dengan metode SFCR semakin menurunkan nilai  $R^2$ , namun pada FGWCR peningkatan  $\delta$  cenderung memperbaiki nilai  $R^2$ .
4. Metode SFCR lebih mampu mempertahankan struktur spasial yang cukup mirip dengan data asli terlihat dari jumlah kluster yang terbentuk.
5. Hasil penelitian ini melibatkan simulasi data spasial berskala besar dan dapat diaplikasikan pada data empiris skala kecamatan bahkan desa. Proses klasterisasi spasial dan regresi secara simultan mengefisienkan waktu dan sumberdaya. Hal ini dapat menjadi masukan metode bagi peneliti atau pemerintah yang membutuhkan pemodelan regresi dengan klasterisasi spasial sekaligus. Salah satu contoh penerapan dapat dilakukan terhadap data kemiskinan atau sebaran penyakit, sehingga pemerintah dapat menentukan kebijakan yang efektif dan efisien untuk setiap daerah dalam maupun antarkluster.

Sintak dari metode yang digunakan penulis dalam artikel ini dapat diakses pada laman <https://github.com/HasanahStk59/sfcr-fgwc>.

## 5 Daftar Pustaka

- [1] A. Djuraidah, *Monograph Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi*. Bogor: IPB Press, 2020.
- [2] S. Cho, D. M. Lambert, S. G. Kim, dan S. Jung, "Extreme coefficients in geographically weighted regression and their effects on mapping," *GIScience Remote Sens.*, vol. 46, no. 3, hal. 273–288, 2009, doi: 10.2747/1548-1603.46.3.273.
- [3] S. Sugawara dan D. Murakami, "Spatially clustered regression," *Spat. Stat.*, vol. 44, hal. 100525, Agu 2021, doi: 10.1016/j.spasta.2021.100525.
- [4] V. K. Malhotra, H. Kaur, dan M. A. Alam, "An Analysis of Fuzzy Clustering Methods," *Int. J. Comput. Appl.*, vol. 94, no. 19, hal. 9–12, 2014.
- [5] B. Tutmez, "Spatial dependence-based fuzzy regression clustering," *Appl. Soft Comput.*,

- 
- vol. 12, no. 1, hal. 1–13, Jan 2012, doi: 10.1016/j.asoc.2011.09.012.
- [6] S. Pramana dan I. H. Pamungkas, “Improvement Method of Fuzzy Geographically Weighted Clustering using Gravitational Search Algorithm,” *J. Ilmu Komput. dan Inf.*, vol. 11, no. 1, hal. 10–16, Feb 2018, doi: 10.21609/jiki.v11i1.580.
- [7] G. Grekousis, “Local fuzzy geographically weighted clustering: a new method for geodemographic segmentation,” *Int. J. Geogr. Inf. Sci.*, vol. 35, no. 1, hal. 152–174, Jan 2021, doi: 10.1080/13658816.2020.1808221.
- [8] I. Made Sumertajaya, M. Nur Aidi, dan W. Nurpadilah, “Mixed Geographically and Temporally Weighted Regression with Cluster in West Java’s Poverty Cases,” *Int. J. Sci. Res. Publ.*, vol. 10, no. 12, hal. 566–572, Des 2020, doi: 10.29322/IJSRP.10.12.2020.p10865.
- [9] M. N. Aidi *et al.*, “Province clustering based on the percentage of communicable disease using the BCBimax biclustering algorithm,” *Geospat. Health*, vol. 18, no. 2, Sep 2023, doi: 10.4081/gh.2023.1202.
- [10] D. Nicholson, O. A. Vanli, S. Jung, dan E. E. Ozguven, “A spatial regression and clustering method for developing place-specific social vulnerability indices using census and social media data,” *Int. J. Disaster Risk Reduct.*, vol. 38, hal. 101224, Agu 2019, doi: 10.1016/j.ijdrr.2019.101224.
- [11] F. Li dan H. Sang, “Spatial Homogeneity Pursuit of Regression Coefficients for Large Datasets,” *J. Am. Stat. Assoc.*, vol. 114, no. 527, hal. 1050–1062, Jul 2019, doi: 10.1080/01621459.2018.1529595.
- [12] I. K. Hasan, Nurwan, Nur Falaq, dan Muhammad Rezky Friesta Payu, “Optimization Fuzzy Geographically Weighted Clustering with Gravitational Search Algorithm for Factors Analysis Associated with Stunting,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 1, hal. 120–128, Feb 2023, doi: 10.29207/resti.v7i1.4508.
- [13] M. Sato-Ilic, “On fuzzy clustering based regression models,” in *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04.*, IEEE, 2004, hal. 216–221 Vol.1. doi: 10.1109/NAFIPS.2004.1336280.
- [14] G. . Mason dan R. . Jacobson, “Fuzzy Geographically Weighted Clustering,” in *Proceedings of the 9th International Conference on Geocomputation*, 2007, hal. 1–7.
- [15] K. Zhou, C. Fu, dan S. Yang, “Fuzziness parameter selection in fuzzy c-means: The perspective of cluster validation,” *Sci. China Inf. Sci.*, vol. 57, no. 11, hal. 1–8, Nov 2014, doi: 10.1007/s11432-014-5146-0.
- [16] D. L. Pham, “Spatial Models for Fuzzy Clustering,” *Comput. Vis. Image Underst.*, vol. 84,

- no. 2, hal. 285–297, Nov 2001, doi: 10.1006/cviu.2001.0951.
- [17] Fotheringham AS, C. Brunsdon, dan C. M, *Geographically Weighted Regression the Analysis of Spatially Varying Relationships*. New York: Wiley, 2002.
- [18] A. A. Suryanto, “Penerapan Metode Mean Absolute Error (Mea) dalam Algoritma Regresi Linear untuk Prediksi Produksi Padi,” *SAINTEKBU*, vol. 11, no. 1, hal. 78–83, Feb 2019, doi: 10.32764/saintekbu.v11i1.298.