

**ANALISIS TINGKAT PENGANGGURAN TERBUKA
DI SUMATERA UTARA TAHUN 2022 MENGGUNAKAN
*GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)***



Disusun Oleh:

Daniella Putri Shalomita	2106631072
Diki Wahyudi	2106709131
Raqi Akbar Robbani	2106652562

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA**

2024

ABSTRAK

Pengangguran masih menjadi salah satu tantangan yang harus dihadapi dalam pembangunan ekonomi regional. Penelitian ini bertujuan untuk menganalisis tingkat pengangguran terbuka di Sumatera Utara tahun 2022 menggunakan pendekatan *Geographically Weighted Regression* (GWR). Metode GWR memungkinkan untuk memahami variabilitas spasial dari faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di wilayah ini. Hasil analisis menunjukkan adanya variasi spasial yang signifikan dalam tingkat pengangguran dan faktor-faktor yang memiliki pengaruh yang berbeda-beda pada tiap lokasi geografisnya. Hal ini memberikan pemahaman yang lebih dalam tentang dinamika pengangguran di tingkat lokal, yang dapat membantu dalam perumusan kebijakan yang lebih efektif dalam upaya mengurangi tingkat pengangguran di Sumatera Utara.

Kata kunci: Pengangguran, *Geographically Weighted Regression*, GWR, Spasial

DAFTAR ISI

ABSTRAK.....	2
DAFTAR ISI.....	3
BAB I	
PENDAHULUAN.....	5
1.1 Latar Belakang.....	5
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
BAB II	
TINJAUAN PUSTAKA.....	7
2.1 Regresi Linier Berganda.....	7
2.1.1. Uji Multikolinieritas.....	8
2.1.2. Uji Normalitas.....	8
2.1.3. Uji Heteroskedastisitas.....	9
2.3 Geographically Weighted Regression.....	10
2.3.1 Pendugaan Parameter Model Geographically Weighted Regression.....	11
2.3.2 Pembobot Spasial.....	11
2.3.3 Uji Parameter Model.....	13
2.3.4 Pengukuran Kualitas Model.....	13
BAB III	
METODE PENELITIAN.....	13
3.1. Data penelitian.....	14
3.2. Jenis Data.....	14
3.3. Variabel Penelitian yang Digunakan.....	14
3.4. Metode Analisis.....	15
BAB IV	
HASIL DAN PEMBAHASAN.....	16
4.1 Visualisasi Variabel Penelitian.....	16
4.2 Analisis Regresi Global.....	16
4.3 Uji Multikolinearitas.....	18
4.4 Uji Asumsi Normalitas.....	19
4.5 Uji Asumsi Heterogenitas Spasial.....	20
4.6 Geographically Weighted Regression (GWR).....	21
4.6.1. Pemilihan Bandwidth Optimum.....	21
4.6.2. Pemilihan Model Terbaik.....	22
4.7 Interpretasi Hasil Estimasi.....	23
4.7.1. Sebaran Kelompok Variabel Signifikan.....	23
4.7.2. Keragaman Spasial Taksiran 1.....	25
4.7.3. Keragaman Spasial Taksiran 2.....	26
4.7.4. Keragaman Spasial Taksiran 3.....	27
4.7.5. Keragaman Spasial Taksiran 4.....	28

4.7.6. Perbandingan Residual Regresi Global dan Regresi Lokal.....	29
BAB IV	
KESIMPULAN.....	31
DAFTAR PUSTAKA.....	32
LAMPIRAN CODE.....	34
LAMPIRAN DATA.....	40
KONTRIBUSI KELOMPOK.....	41

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sebagai negara berkembang, Indonesia masih memiliki berbagai permasalahan, salah satunya adalah kemiskinan. Kemiskinan merupakan masalah kompleks yang dipengaruhi oleh berbagai faktor yang saling berkaitan. Menurut I Komang, tingkat pengangguran terbuka dapat menjadi salah satu faktor yang berkaitan erat dengan kemiskinan. Di Indonesia, tingkat pengangguran terbuka menjadi perhatian utama pemerintah dalam upaya mencapai pertumbuhan ekonomi yang inklusif dan berkelanjutan. Salah satunya adalah Provinsi Sumatera Utara yang berkontribusi signifikan terhadap ekonomi nasional, tidak terkecuali dari tantangan tingkat pengangguran.

Terlebih lagi, tahun 2022 dunia dihadapkan pada berbagai tantangan, termasuk pandemi Covid-19 yang sangat berdampak terhadap perekonomian, tak terkecuali Indonesia. Pandemi Covid-19 telah menyebabkan penurunan aktivitas ekonomi, salah satunya yang sangat terasa adalah pengurangan lapangan pekerjaan. Pengurangan lapangan pekerjaan ini tentunya akan meningkatkan angka pengangguran karena ketidakseimbangan antara jumlah lapangan pekerjaan dengan bertambahnya angkatan kerja.

Dalam penelitian ini, penggunaan metode analisis *Geographically Weighted Regression* (GWR) dinilai dapat memodelkan kasus tingkat pengangguran terbuka di daerah Sumatera Utara karena metode ini memungkinkan untuk memahami variabilitas spasial dalam data dan mengidentifikasi pola-pola lokal yang tidak terdeteksi dalam analisis regresi linier berganda. Dengan menerapkan metode GWR, diharapkan nantinya dapat diidentifikasi faktor-faktor spesifik yang berkaitan dengan tingkat pengangguran terbuka di Sumatera Utara yang berbeda-beda di setiap wilayahnya.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apa model terbaik untuk memodelkan kasus tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022?
2. Bagaimana hasil pemodelan GWR pada kasus tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022?
3. Apa saja faktor-faktor yang signifikan berpengaruh pada tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui model terbaik untuk memodelkan kasus pengangguran terbuka di Sumatera Utara pada tahun 2022.
2. Untuk mengetahui hasil pemodelan GWR pada kasus tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022.
3. Untuk mengetahui faktor-faktor yang signifikan berpengaruh pada tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022.

BAB II

TINJAUAN PUSTAKA

2.1 Regresi Linier Berganda

Regresi linier berganda adalah model yang menunjukkan hubungan antara dua atau lebih variabel independen (X) dengan satu variabel dependen (Y). Uji regresi linier berganda bertujuan untuk menentukan nilai-nilai koefisien β yang memberikan estimasi terbaik untuk variabel dependen Y berdasarkan variabel-variabel independen yang diberikan. Selain itu juga dapat mengetahui arah hubungan variabel dependen dengan variabel-variabel independennya.

Secara sistematis, persamaan untuk n observasi dengan variabel independen sebanyak k adalah sebagai berikut.

$$Y_i = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \dots + \beta_k X_{ik} + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (2.1)$$

dengan:

- Y_i : nilai observasi variabel dependen ke- i
- β_0 : nilai *intercept* model regresi
- β_k : koefisien regresi variabel independen ke- k
- X_{ik} : nilai observasi variabel independen ke- k pada pengamatan ke- i
- ε_i : *error* pada pengamatan ke- i

- Hipotesis untuk uji simultan (uji F) adalah sebagai berikut:
 - $H_0 : \beta = 0$ (variabel X tidak berpengaruh signifikan)
 - $H_1 : \beta \neq 0$ (setidaknya terdapat satu variabel yang berpengaruh)
 - Hipotesis untuk uji parsial (uji t) adalah sebagai berikut:
 - $H_0 : \beta_i = 0$ (variabel X tidak berpengaruh signifikan)
 - $H_1 : \beta_i \neq 0$ (setidaknya terdapat satu variabel yang berpengaruh)
- dengan $i = 0, 1, 2, \dots, n$*

Dalam melakukan pemodelan regresi, terdapat beberapa asumsi yang harus dipenuhi diantaranya uji multikolinieritas dan uji asumsi residual yaitu uji normalitas, uji homoskedastisitas, dan uji autokorelasi.

2.1.1. Uji Multikolinieritas

Uji multikolinieritas ditujukan untuk menguji adanya hubungan linier antara beberapa variabel independen dalam suatu model regresi linier berganda. Model regresi yang baik adalah model yang variabel independennya tidak memiliki korelasi. Uji ini menggunakan nilai *Tolerance* (TOL) dan *Variance Inflation Factor* (VIF). Apabila nilai TOL < 0.1 atau nilai VIF > 10, maka terdapat masalah multikolinieritas. Nilai TOL dan VIF adalah sebagai berikut.[3]

$$VIF = \frac{1}{R^2_{yj1}} \quad (2.2)$$

$$TOL = \frac{1}{VIF} = (1 - R^2_{yj1}) \quad (2.3)$$

dengan
$$R^2_{yj1} = \frac{r^2_{yj} + r^2_{y1} + 2r_{yj}r_{y1}r_{j1}}{1 - r^2_{j1}} \quad (2.4)$$

2.1.2. Uji Normalitas

Uji normalitas digunakan untuk menguji apakah residual terdistribusi secara normal atau tidak. Uji normalitas dapat dilakukan dengan uji *Jarque-Berra*, *Anderson Darling*, *Kolmogorov-Smirnov*, dan *Skewness-Kurtosis*. Berikut merupakan uji normalitas dengan *Kolmogorov-Smirnov*:

- Hipotesis

H_0 : Residual berdistribusi normal

H_1 : Residual tidak berdistribusi normal

- Tingkat Signifikansi

$\alpha = 0,05$

- Statistik Uji *Kolmogorov-Smirnov*

$$D = \max |F_0 - S_n(x_i)|, i = 1, 2, \dots, N$$

dengan:

$F_0(x)$ = fungsi distribusi kumulatif yang ditentukan

$S_n(x)$ = frekuensi kumulatif yang diobservasi dari sampel acak dengan N observasi

- Daerah Kritis

Tolak H_0 jika $D_{hitung} > D_{tabel}$ atau $p\text{-value} < \alpha$

2.1.3. Uji Heteroskedastisitas

Uji heteroskedastisitas yang dilakukan bertujuan untuk menguji apakah dalam model regresi tersebut memiliki varian yang konstan dari residual atau error antara satu pengamatan ke pengamatan yang lain. Apabila terjadi heteroskedastisitas dalam model regresi, maka estimator yang diperoleh tidak efisien. Terdapat beberapa uji dalam mendeteksi heteroskedastisitas, yaitu uji *Park* dan uji *Breusch Pagan Godfrey*. [7]. Langkah-langkah dalam melakukan uji *Breusch Pagan Godfrey* adalah sebagai berikut.

- Lakukan regresi OLS persamaan

$$Y_i = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \dots + \beta_k X_{ik} + \varepsilon_i \text{ sehingga didapat nilai}$$

$$\text{residualnya dan mencari } \sigma^2 = \frac{\sum \hat{e}_i^2}{n} \quad (2.5)$$

- Mencari p_i yang didefinisikan sebagai: $p_i = \frac{\hat{e}_i^2}{\sigma^2} \quad (2.6)$

- Regresi p_i terhadap variabel Z sebagai berikut: $p_i = \alpha_0 + \alpha_1 Z_i = v_i$

- Dapatkan ESS (*explained sum of square*) dan kemudian dapatkan

$$\phi = \frac{1}{2} (ESS). \text{ Jika residual berdistribusi normal, maka akan}$$

mengikuti distribusi Chi-Square (χ^2) sebagai berikut:

$$\phi = \frac{1}{2} (ESS) \approx \chi_{df}^2$$

2.2 Heterogenitas Spasial

Data spasial merupakan data yang memuat informasi geografis terkait lokasi observasi. Heterogenitas spasial merupakan keadaan dimana variabel independen memberikan respon tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian (Tizona, Goejantoro, Wasono, 2017). Data spasial dapat diuji akan

keberadaan heterogenitas spasial. Salah satu cara untuk mengujinya adalah dengan menggunakan statistik uji breusch-pagan.

Hipotesis

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma^2$ (tidak terjadi heterogenitas antar wilayah)

$H_1: \sigma_i^2 \neq \sigma^2$, setidaknya untuk satu i (terjadi heterogenitas antar wilayah)

Statistik Uji:

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f + \left(\frac{1}{T} \right) \left(\frac{\varepsilon^T W \varepsilon}{\sigma^2} \right)^2, \quad (2.7)$$

$$\text{dimana vektor } f \text{ adalah } f_i = \left(\frac{s_i^2}{\sigma^2} - 1 \right) \quad (2.8)$$

$$\text{dan } T = \text{tr}[W^T W + W^2]. \quad (2.9)$$

Keterangan:

BP : Statistik uji *Breusch-Pagan*

ε : vektor residual ε_i berukuran $n \times 1$

ε_i : residual ke- i , $i = 1, 2, 3, \dots, n$

f^T : (f_1, f_2, \dots, f_n)

Z : matriks X telah distandarisasi, $n \times (p + 1)$

σ^2 : ragam residual ε_i

W : matriks pembobot W_i berukuran $n \times n$

n : banyaknya pengamatan

p : banyaknya variabel prediktor

Aturan keputusan dari uji ini adalah tolak H_0 apabila statistik uji $BP > \chi^2_{\left(\frac{\alpha}{2}, p+1\right)}$.

2.3 Geographically Weighted Regression

Geographically Weighted Regression (GWR) adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi observasi, sehingga di setiap lokasi mempunyai nilai parameter model GWR yang berbeda-beda. Parameter yang dihasilkan pada model GWR akan berbeda-beda pada masing-masing lokasi, sehingga terdapat sebanyak $n \times (p + 1)$ parameter yang harus diestimasi, dimana n adalah jumlah lokasi pengamatan dan p adalah jumlah peubah bebas. Analisis data

spasial dilakukan menggunakan metode GWR ketika dideteksi terdapat efek heterogenitas spasial.

Bentuk umum model *Geographically Weighted Regression* (GWR).

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)X_{ik} + \varepsilon_i, i = 1, 2, \dots, n \quad (2.10)$$

Keterangan:

- Y_i : nilai variabel dependen pada lokasi ke i
 X_{ik} : nilai variabel independen ke k pada lokasi ke i
 (u_i, v_i) : koordinat lokasi pada lokasi ke i
 $\beta_k(u_i, v_i)$: nilai parameter ke k pada lokasi ke i
 $\beta_0(u_i, v_i)$: nilai parameter intercept pada lokasi ke i
 ε_i : nilai error pada lokasi ke i

2.3.1 Pendugaan Parameter Model *Geographically Weighted Regression*

Pada GWR, sebuah observasi diberi bobot sesuai dengan perkiraan lokasi i , sehingga nilai observasi bervariasi berdasarkan lokasi i . Sederhananya, sebuah observasi yang lebih dekat dari lokasi i memiliki bobot yang lebih besar dibandingkan dengan observasi yang jauh dari lokasi i . Dengan begitu, didapatkan persamaan penduga parameter sebagai berikut.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (2.11)$$

Keterangan:

- $\hat{\beta}(u_i, v_i)$: matriks penduga nilai parameter pada lokasi ke i
 $W(u_i, v_i)$: matriks pembobot untuk lokasi ke i

2.3.2 Pembobot Spasial

Pada 2.7.1 sebelumnya telah disebut tentang pembobot $W(u_i, v_i)$. Matriks pembobot spasial ini berisi elemen yaitu $w_k(u_i, v_i)$ dengan ukuran matriksnya $n \times n$.

$$W(u_i, v_i) = \begin{bmatrix} w_1(u_i, v_i) & 0 & \dots & 0 & 0 \\ 0 & w_2(u_i, v_i) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & w_{n-1}(u_i, v_i) & 0 \\ 0 & 0 & \dots & 0 & w_n(u_i, v_i) \end{bmatrix}$$

Nilai dari setiap pembobot spasial ($w_1(u_1, v_1)$) didapatkan dari fungsi pembobot spasial. Fungsi pembobot spasial memberikan nilai 0 sampai 1 berdasarkan jarak antar titik berdasarkan model dengan titik observasi. Nilai semakin mendekati 1 akan didapatkan ketika titik model makin kecil jaraknya ke titik observasi.

Dengan fungsi pembobot spasial K didapatkan persamaan berikut.

$$w_i(u_i, v_i) = K(d_{ij}), \quad (2.12)$$

$$\text{dimana } d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (2.13)$$

Pada penelitian ini digunakan tiga fungsi pembobot spasial yaitu Fungsi Kernel Gaussian, Fungsi Kernel Bisquare, dan Fungsi Kernell Tricube yang memiliki persamaan sebagai berikut.

- Fungsi Kernel Gaussian:

$$K_G(d_{ij}) = \exp\left[-\frac{1}{2} \left(\frac{d_{ij}}{b}\right)^2\right] \quad (2.14)$$

- Fungsi Kernel Bisquare:

$$K_B(d_{ij}) = \left(1 - \frac{1}{2} \left(\frac{d_{ij}}{b}\right)^2\right)^2, \quad (2.15)$$

jika $d_{ij} > b$, maka $K_B(d_{ij}) = 0$

Keterangan:

d_{ij} : jarak antara titik berdasarkan model (i) dengan titik observasi (j)

b : nilai *bandwidth* optimum

- Fungsi Kernel Tricube:

$$K_T(d_{ij}) = \left(1 - \frac{1}{2} \left(\frac{d_{ij}}{2.18452}\right)^3\right)^3 \quad (2.16)$$

Penentuan nilai *bandwidth* optimum (b) ditentukan dengan menghitung nilai cross-validation terkecil menggunakan persamaan di bawah. Nilai b akan didapatkan ketika CV mencapai nilai minimum.

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2 \quad (2.17)$$

Keterangan:

$y_{\neq i}$: nilai estimasi untuk variabel respon selain ke- i .

Fungsi pembobot spasial yang baik memenuhi sifat-sifat berikut.

- $K(0) = 1$
- $\lim_{d_{ij} \rightarrow 0} K(d_{ij}) = 0$
- $K(d_{ij})$ adalah fungsi monoton turun

2.3.3 Uji Parameter Model

Model yang telah dibangun akan diuji parameternya yang dapat dilakukan secara multivariat atau univariat.

- Uji Parsial

$$H_0: \beta_k(u_i, v_i) = 0$$

$$H_1: \text{Setidaknya satu } \beta_p(u_i, v_i) \neq 0$$

- Uji Serentak

$$H_0: \beta_k(u_i, v_i) = 0$$

$$H_1: \beta_k(u_i, v_i) \neq 0$$

2.3.4 Pengukuran Kualitas Model

Pada penelitian ini hanya digunakan satu metode untuk menentukan model yang terbaik, yaitu AIC (*Akaike Information Criterion*). AIC adalah ukuran relatif kebaikan fit dari model statistik secara umum dengan memperkirakan kualitas masing-masing model, relatif terhadap masing-masing model potensial lainnya. Nilai AIC yang lebih kecil mengartikan bahwa model tersebut lebih baik. Berikut cara menentukan nilai AIC.

$$AIC = 2p - 2 \ln(L), \quad (2.18)$$

dimana $L = L(\hat{\theta})$ (nilai maksimum fungsi likelihood).

BAB III

METODE PENELITIAN

3.1. Data penelitian

Data penelitian yang digunakan merupakan data sekunder dari website Badan Pusat Statistik (BPS) Provinsi Sumatera Utara.

3.2. Jenis Data

Jenis data yang digunakan merupakan data spasial yang memiliki variabel koordinat.

3.3. Variabel Penelitian yang Digunakan

Dalam melakukan penelitian terkait analisis Tingkat Pengangguran Terbuka (TPK) di Sumatera Utara tahun 2022, variabel-variabel yang digunakan tertera pada Tabel 1 sebagai berikut.

Tabel 3.1. Variabel Penelitian

No.	Kategori	Variabel	Jenis Variabel	Penjelasan Variabel
1	Variabel Independen	Persentase Penduduk Miskin (PPM); X_1	Numerik	Persentase penduduk yang berada dibawah Garis Kemiskinan (GK).
2		<i>Dependency Ratio</i> (DR); X_2	Numerik	Angka yang menyatakan perbandingan antara banyaknya penduduk usia nonproduktif (penduduk di bawah 15 tahun dan penduduk diatas 65 tahun) dengan banyaknya penduduk usia produktif (penduduk usia 15–64 tahun).
3		Gini Rasio (GR); X_3	Numerik	Alat mengukur derajat ketidakmerataan distribusi penduduk; dimana jika $<0,4$ maka tingkat ketimpangan rendah; jika $0,4-0,5$, maka tingkat ketimpangan sedang; jika $>0,5$, maka tingkat ketimpangan tinggi.

4		Laju Pertumbuhan Penduduk (LPP); X_4	Numerik	Angka yang menunjukkan rata-rata tingkat pertumbuhan penduduk per tahun dalam jangka waktu tertentu.
5	Variabel Dependen	Tingkat Pengangguran Terbuka (TPK); Y	Numerik	Tingkat pengangguran terbuka adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja

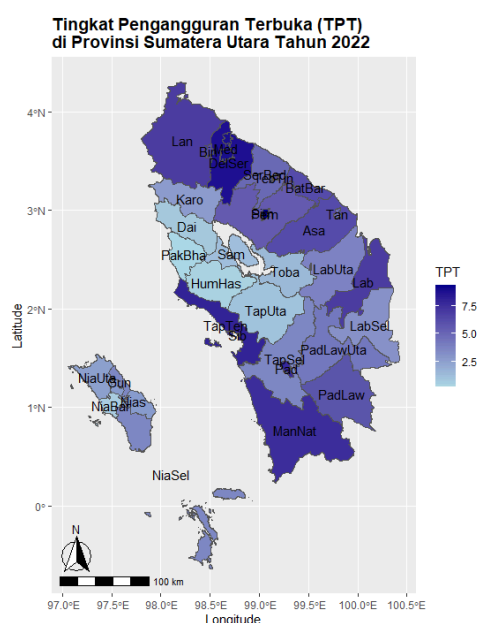
3.4. Metode Analisis

Dalam penelitian ini, analisis dilakukan dengan menggunakan Regresi Linier Berganda dan juga *Geographically Weighted Regression* (GWR).

HASIL DAN PEMBAHASAN

4.1 Visualisasi Variabel Penelitian

Tingkat pengangguran kabupaten/kota di setiap kabupaten/kota beragam dari yang terendah di Kabupaten Pakpak Bharat sebesar 0.26 sampai tertinggi di Kabupaten Pematangsiantar sebesar 9.36. Data sebaran tingkat pengangguran terbuka di Sumatera Utara tahun 2022 dipetakan pada Gambar 4.1.



Gambar 4.1. Peta Tingkat Pengangguran Terbuka Tahun 2022 di Sumatera Utara

Berdasarkan gambar di atas, terlihat bahwa terdapat pola mengelompok pada tingkat pengangguran terbuka yang rendah di Sumatera Utara.

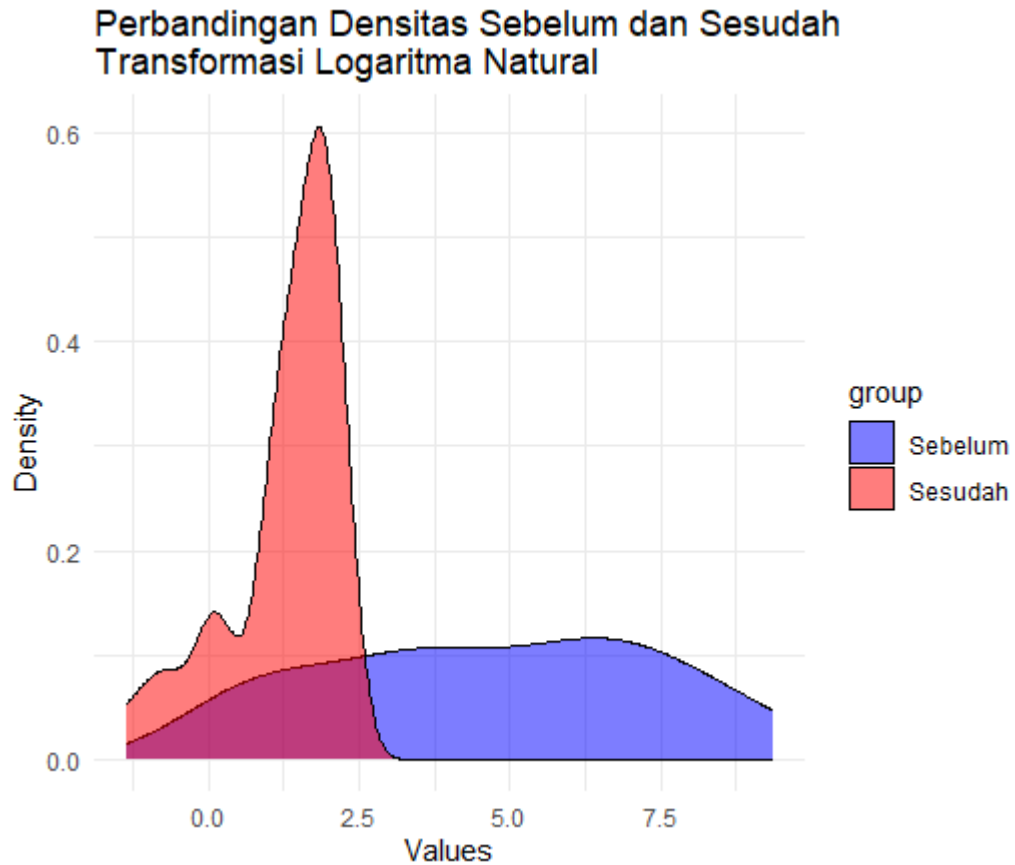
4.2 Analisis Regresi Global

Pemodelan regresi global ini bertujuan untuk melakukan analisis awal pada data tingkat pengangguran terbuka. Pendugaan parameter model regresi global dilakukan dengan metode *Ordinary Least Square* (OLS). Berikut ini merupakan model regresi yang digunakan untuk data ini.

$$\ln(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i \quad (4.1)$$

Dalam persamaan di atas, dilakukan transformasi logaritma natural pada variabel Y untuk menyesuaikan dengan tujuan dari penelitian ini, yaitu melakukan pemodelan

GWR. Jika tidak ditransformasi, variabel Y tidak akan mengalami heteroskedastisitas spasial sehingga tidak dapat dilakukan pemodelan GWR. Berikut ini plot perbandingan densitas variabel Y sebelum dan sesudah dilakukan transformasi.



Gambar 4.2. Perbandingan Densitas Variabel Y Sebelum dan Sesudah Transformasi Logaritma Natural

Terlihat bahwa setelah transformasi, densitas variabel Y menjadi memiliki puncak dibandingkan sebelum transformasi. Kedua densitas tersebut tidak terlalu mirip dengan distribusi normal.

Tabel 4.1 Pendugaan Parameter dengan Metode OLS

Variabel	Koefisien	Nilai t	Std. Error	$p - value$
<i>Intercept</i>	9.5074	1.91227	4.972	0.00003
X_1 (PPM)	-0.07674	0.02746	-2.794	0.0092800
X_2 (DR)	-0.15313	0.03079	-4.973	0.0000299
X_3 (GR)	1.42532	2.95534	0.482	0.63335
X_4 (LPP)	-0.77195	0.28754	-2.685	0.01206

Tabel 4.1 adalah ringkasan hasil pendugaan parameter dari model regresi global. Dari tabel tersebut, taksiran model regresi global untuk data ini adalah:

$$\ln(Y_i) = 9.5074 - 0.07674X_{1i} - 0.15313X_{2i} + 1.42532X_{3i} - 0.77195X_{4i} \quad (4.2)$$

Hasil uji global F didapatkan bahwa nilai $F = 9.29$ dan $p - value = 6.573 \times 10^{-5} < \alpha = 0.05$. Dari hasil tersebut, didapatkan bahwa H_0 ditolak karena $p - value < \alpha$. Berarti, setidaknya terdapat satu variabel independen yang memberikan pengaruh signifikan terhadap variabel dependen.

Berdasarkan tabel 4.1, menggunakan uji parsial t , didapatkan bahwa hanya variabel X_3 (Gini Rasio) yang tidak memberikan pengaruh signifikan terhadap variabel dependen (Tingkat Pengangguran Terbuka) pada tingkat signifikansi 5%. Nilai AIC model yang diperoleh adalah sebesar 72.98903. Koefisien determinasi (R^2) model adalah 0.5703. Berarti, hanya 57.03% variabilitas dari variabel dependen yang dapat dijelaskan oleh variabel independen pada model (4.2).

4.3 Uji Multikolinearitas

Pengujian asumsi multikolinieritas dapat dilihat berdasarkan nilai *Variance Inflation Factor* (VIF). Multikolinearitas terjadi saat VIF bernilai lebih dari 10. Berikut merupakan nilai VIF setiap variabel independen.

```
> vif(reg)
PPM_2022 DR_2022 GR_2022 LPP_2022
1.092928 1.070644 1.109938 1.040939
```

Gambar 4.3 Output VIF Model Regresi

Berdasarkan *output* di atas, tidak ada nilai VIF yang lebih dari 10. Jadi, tidak ada masalah multikolinieritas pada model ini.

4.4 Uji Asumsi Normalitas

- Hipotesis:
 H_0 : Residual Mengikuti Distribusi Normal
 H_1 : Residual Tidak Mengikuti Distribusi Normal
- Tingkat Signifikansi:
 $\alpha = 0.05$
- Statistik Uji:

```
> shapiro.test(reg$residuals)

      shapiro-wilk normality test

data:  reg$residuals
W = 0.98223, p-value = 0.8507

> ks.test(reg$residuals, "pnorm")

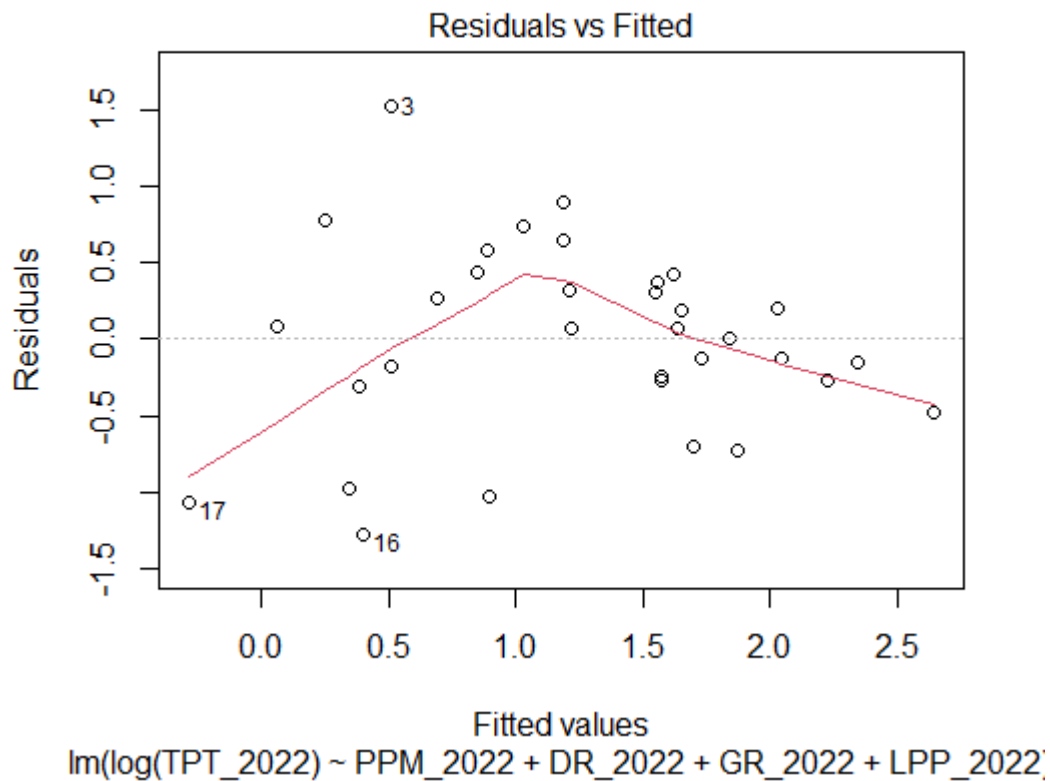
      Exact one-sample kolmogorov-smirnov test

data:  reg$residuals
D = 0.16467, p-value = 0.2987
alternative hypothesis: two-sided
```

Gambar 4.4 Output Uji Kolmogorov-Smirnov 1 Sampel dan Uji Shapiro-Wilk

- Aturan Keputusan dan Kesimpulan:
Tolak H_0 jika $p - value < \alpha$. Karena kedua uji memberikan hasil $p - value > 0.05$, maka kedua uji menghasilkan hasil yang sama yaitu H_0 diterima. Jadi, residual model regresi mengikuti distribusi normal.

4.5 Uji Asumsi Heterogenitas Spasial



Gambar 4.5 Plot Residual Dibandingkan dengan Taksiran Nilai \hat{Y}

Dari gambar 4.5, terlihat bahwa terdapat pola yang tidak *random* pada titik-titik di plot. Karena data yang dianalisis merupakan data spasial, data dengan *latitude* dan *longitude*, maka dapat dicurigai pola tidak *random* tersebut merupakan heterogenisasi variansi residual yang disebabkan oleh adanya pengaruh spasial. Oleh karena itu, dilakukan uji heterogenitas spasial dengan menggunakan uji Breusch-Pagan untuk mendeteksi adanya heterogenitas spasial.

- Hipotesis:
 H_0 : Tidak terdapat heterogenitas spasial
 H_1 : terdapat heterogenitas spasial
- Tingkat Signifikansi:
 $\alpha = 0.05$
- Statistik Uji:

```
> bptest(reg)

studentized Breusch-Pagan test

data:  reg
BP = 10.013, df = 4, p-value = 0.0402
```

Gambar 4.6 *Output Uji Breusch-Pagan*

- Aturan Keputusan dan Kesimpulan:
Tolak H_0 jika $p - value < \alpha$. Karena uji memberikan hasil $p - value = 0.0402 < \alpha = 0.05$, maka H_0 ditolak. Jadi, terdapat heterogenitas spasial pada data. Oleh karena itu, pemodelan regresi global dapat dilanjutkan ke pemodelan regresi lokal dengan menggunakan Geographically Weighted Regression (GWR).

4.6 *Geographically Weighted Regression (GWR)*

4.6.1. *Pemilihan Bandwidth Optimum*

Bandwidth optimum digunakan sebagai parameter fungsi pembobot spasial. Pada bagian ini, akan dilakukan optimasi *bandwidth* dengan tiga fungsi kernel, yaitu Gaussian, Bisquare, Tricube, dan kombinasinya untuk *adaptive bandwidth* dan *fixed bandwidth*. Untuk fungsi pembobot Kernel Gaussian, diperoleh *fixed bandwidth* sebesar 1.108295. Fungsi Kernel Gaussian yang digunakan untuk membuat matriks pembobot adalah

$$K_G(d_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{1.108295}\right)^2\right] \quad (4.3)$$

Bandwidth 1.108295 merupakan jarak dalam satuan derajat dalam sistem koordinat geografi. Interpretasi dari nilai *bandwidth* tersebut adalah kabupaten/kota yang berjarak tidak lebih *bandwidth* 1.108295 dari kabupaten/kota yang diamati akan mendapatkan bobot yang secara signifikan lebih besar dibanding dengan kabupaten/kota yang berjarak lebih dari *bandwidth* 1.108295 sehingga memiliki pengaruh yang lebih besar terhadap pendugaan parameter model GWR.

Untuk fungsi pembobot Kernel Bisquare, *fixed bandwidth* optimum yang diperoleh sebesar 3.331662 sehingga fungsi pembobot Kernel Bisquare yang digunakan adalah

$$K_B(d_{ij}) = \left(1 - \frac{1}{2}\left(\frac{d_{ij}}{3.331662}\right)^2\right)^2 \quad (4.4)$$

Interpretasinya adalah kabupaten/kota yang berjarak lebih dari *bandwidth* 3.331662 dari kabupaten/kota yang diamati akan mendapatkan bobot nol sehingga tidak mempengaruhi nilai pendugaan parameter model GWR.

Untuk fungsi pembobot Kernel Tricube, *fixed bandwidth* optimum yang didapatkan adalah 3.428924 sehingga fungsi pembobot Kernel Tricube yang digunakan adalah

$$K_T(d_{ij}) = (1 - \frac{1}{2}(\frac{d_{ij}}{3.428924})^3)^3 \quad (4.5)$$

Interpretasinya adalah kabupaten/kota yang berjarak lebih dari *bandwidth* 3.428924 dari kabupaten/kota yang diamati akan mendapatkan bobot nol sehingga tidak mempengaruhi nilai pendugaan parameter model GWR.

Untuk Gaussian kernel dengan *adaptive banddwith* tidak dilakukan estimasi karena persamaan lokal pada observasi ke-17 dan 18 tidak *full* rank. Jadi, hanya ada lima model GWR yang akan diestimasi.

4.6.2. Pemilihan Model Terbaik

Setelah mendapatkan nilai *bandwidth* yang optimal, akan dipilih model terbaik berdasarkan kriteria AIC dan R^2 dari masing-masing model. Model terbaik adalah model dengan skor AIC terkecil serta R^2 terbesar.

Tabel 4.2 Nilai AIC dan R^2 Model

Model	AIC	R^2
Regresi Linier Global	72.98903	0.5703
GWR Gaussian (<i>Fixed</i>)	53.14388	0.7422885
GWR Bisquare (<i>Fixed</i>)	64.41965	0.5954222
GWR Bisquare (<i>Adaptive</i>)	46.21338	0.804984
GWR Tricube (<i>Fixed</i>)	64.50638	0.5940667
GWR Tricube (<i>Adaptive</i>)	-30.80109	0.9878199

Dari tabel 4.2, didapatkan bahwa model GWR Tricube (*adaptive*) merupakan model terbaik karena memiliki nilai AIC terkecil dan R^2 terbesar. Berikut ini merupakan hasil estimasi dari model GWR Tricube (*adaptive*).

Tabel 4.3 Hasil Pemodelan *Geographically Weighted Regression* Kernel Tricube dengan *Adaptive Bandwidth*

Koefisien	Minimum	Median	Maksimum
β_{0i} (Intercept)	-8.95021	8.912389	17.24528
β_{1i} (PPM)	-0.23229	0.036909	0.287722
β_{2i} (DR)	-0.27208	-0.16745	0.067638
β_{3i} (GR)	-13.3329	2.786317	29.30364
β_{4i} (LPP)	-1.23127	-0.61716	2.341313

Dari estimasi parameter yang didapatkan pada Tabel 4.3, didapatkan hubungan yang berbeda-beda antara variabel independen terhadap variabel dependen. Didapatkan estimasi parameter lokal dengan nilai *range* tertinggi adalah parameter dari variabel gini rasio (X_4). Artinya, variabilitas efek dari gini rasio terhadap tingkat pengangguran terbuka (Y) tinggi.

4.7 Interpretasi Hasil Estimasi

Pada bagian sebelumnya telah diketahui bahwa model terbaik untuk menganalisis data pada penelitian ini adalah model GWR dengan fungsi pembobot kernel Tricube dengan *adaptive bandwidth*. Pada bagian ini, akan dilakukan interpretasi dari masing-masing hasil estimasi parameter model regresi lokal GWR.

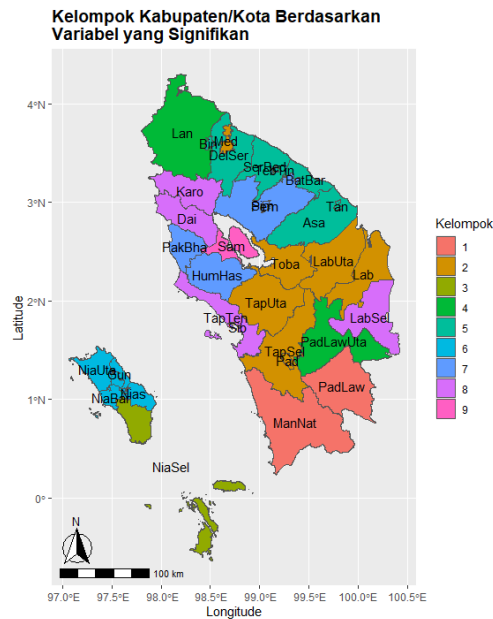
4.7.1. Sebaran Kelompok Variabel Signifikan

Terdapat sembilan kelompok lokasi berdasarkan variabel independen yang signifikan seperti yang ditampilkan pada Tabel 4.4.

Tabel 4.4 Hasil Pemodelan *Geographically Weighted Regression* Kernel Tricube dengan *Adaptive Bandwidth*

Kelompok	Variabel Signifikan	Jumlah Kab/Kota	Nama Kab/Kota
1	Tidak ada	2	Mandailing Natal, Padang Lawas
2	X_2	7	Labuhanbatu, Labuhanbatu Utara, Medan, Padangsidempuan, Tapanuli Selatan, Tapanuli Utara, Toba
3	X_3	1	Nias Selatan
4	X_4	2	Langkat, Padang Lawas Utara
5	X_2, X_4	7	Asahan, Batu Bara, Binjai, Deli Serdang, Serdang Bedagai, Tanjungbalai, Tebing Tinggi
6	X_3, X_4	4	Gunungsitoli, Nias, Nias Barat, Nias Utara
7	X_1, X_2, X_4	4	Humbang Hasundutan, Pakpak Bharat, Pematangsiantar, Simalungun
8	X_2, X_3, X_4	5	Dairi, Karo, Labuhanbatu Selatan, Sibolga, Tapanuli Tengah
9	X_1, X_2, X_3, X_4	1	Samosir

Dari keempat peubah bebas yang dianalisis, X_2 (*dependency ratio*) menjadi variabel yang paling sering menjadi variabel signifikan. Kelompok dengan anggota terbanyak adalah kelompok 2 dan kelompok 5, dengan anggota kelompok sebanyak 7. Berikut ini peta dari kelompok yang terbentuk.

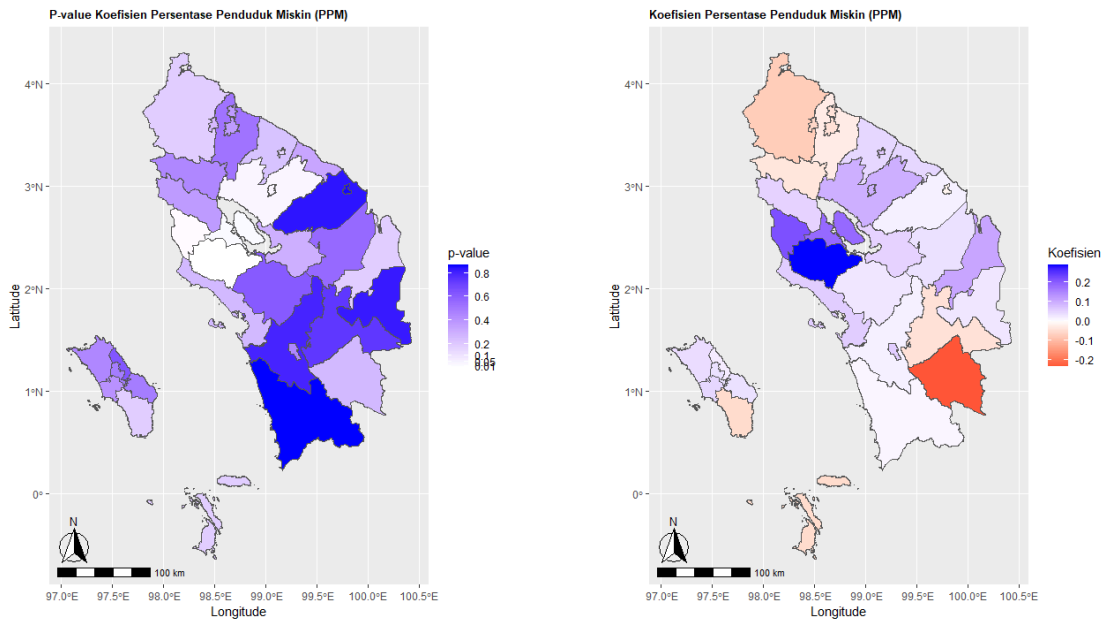


Gambar 4.7 Peta Kelompok Kabupaten/Kota Berdasarkan Variabel yang Signifikan

Berdasarkan plot di atas, terlihat bahwa kelompok yang terbentuk cenderung berdasarkan wilayah yang terdekat. Contohnya, kabupaten/kota yang sama sekali tidak memiliki variabel yang signifikan pada persamaan lokal, yaitu Mandailing Natal dan Padang Lawas, membentuk kelompok di wilayah selatan. Sedangkan, kelompok 5, yaitu kabupaten/kota dengan variabel X_2 dan X_4 yang signifikan, membentuk kelompok di wilayah utara.

4.7.2. Keragaman Spasial Taksiran β_1

Gambar 4.8 menunjukkan sebaran nilai taksiran β_1 , beserta p - value-nya, yang menggambarkan pengaruh rata-rata persentase penduduk miskin (X_1) terhadap tingkat pengangguran terbuka (Y) di setiap kabupaten/kota di Sumatera Utara.

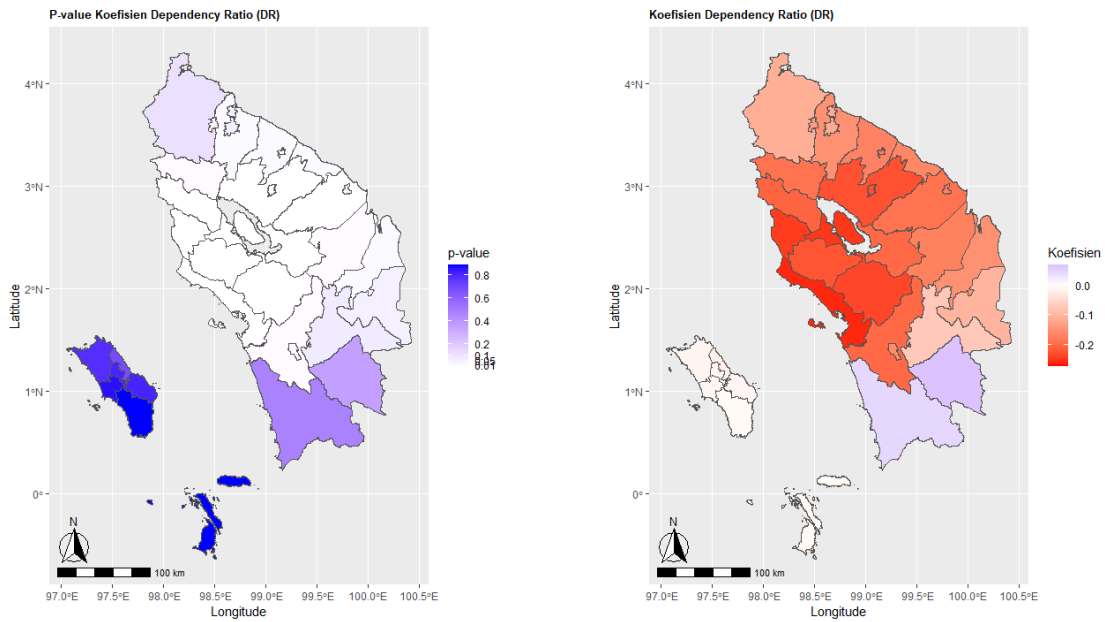


Gambar 4.8 Peta Keragaman Spasial Taksiran β_1

Di beberapa kabupaten/kota yang saling berdekatan, dapat dilihat bahwa nilai taksiran parameternya hampir sama. Warna biru atau ungu (plot kanan) menandakan rata-rata persentase penduduk miskin berpengaruh secara positif terhadap tingkat pengangguran terbuka di kabupaten/kota, yang cenderung berada di bagian tengah wilayah Provinsi Sumatera Utara.

4.7.3. Keragaman Spasial Taksiran β_2

Gambar 4.9 menunjukkan sebaran nilai taksiran β_2 , beserta p - $value$ -nya, yang menggambarkan pengaruh rata-rata *dependency ratio* (X_2) terhadap tingkat pengangguran terbuka (Y) di setiap kabupaten/kota di Sumatera Utara.

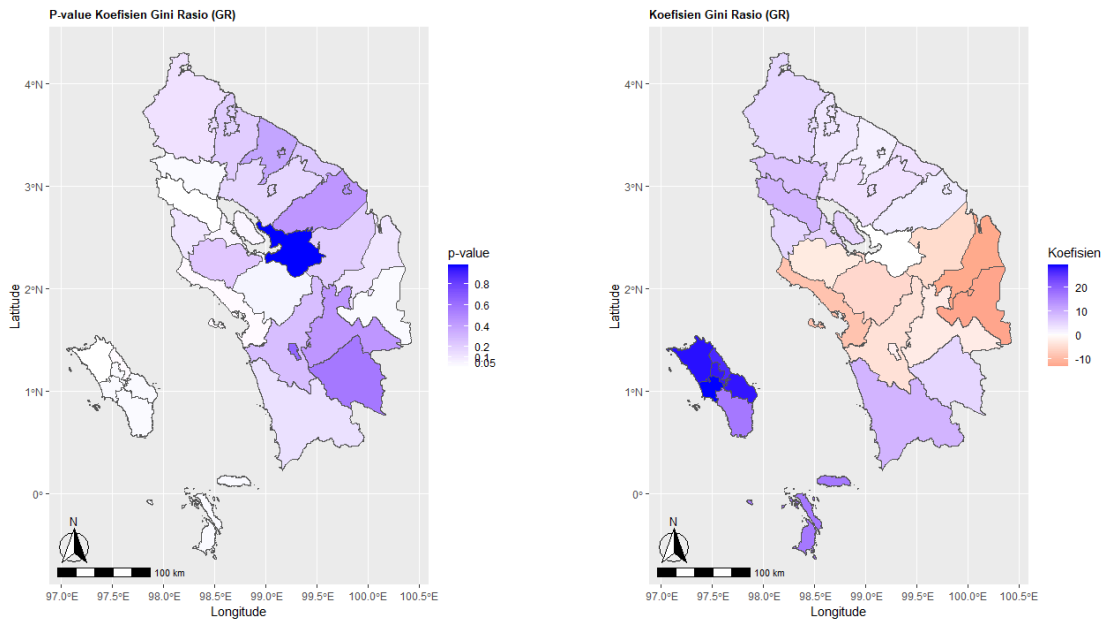


Gambar 4.9 Peta Keragaman Spasial Taksiran β_2

Di beberapa kabupaten/kota yang saling berdekatan, dapat dilihat bahwa nilai taksiran parameternya hampir sama. Warna merah (plot kanan) menandakan rata-rata *dependency ratio* berpengaruh secara negatif terhadap tingkat pengangguran terbuka di kabupaten/kota wilayah Provinsi Sumatera Utara. Untuk efek *dependency ratio*, efeknya secara umum berbanding terbalik dengan tingkat pengangguran terbuka. Wilayah yang memiliki pengaruh *dependency ratio* yang signifikan cenderung berada di wilayah tengah pulau Sumatera Utara (berdasarkan plot kiri).

4.7.4. Keragaman Spasial Taksiran β_3

Gambar 4.10 menunjukkan sebaran nilai taksiran β_3 , beserta *p-value*-nya, yang menggambarkan pengaruh rata-rata gini rasio (X_3) terhadap tingkat pengangguran terbuka (Y) di setiap kabupaten/kota di Sumatera Utara.

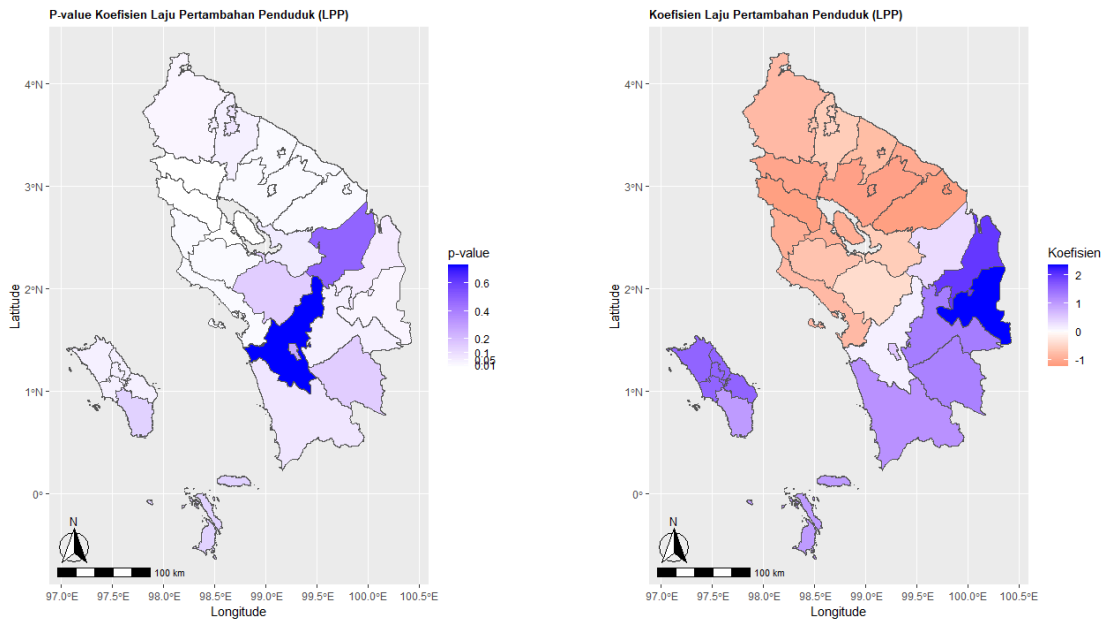


Gambar 4.10 Peta Keragaman Spasial Taksiran β_3

Di beberapa kabupaten/kota yang saling berdekatan, dapat dilihat bahwa nilai taksiran parameternya hampir sama. Warna biru muda (plot kanan) menandakan rata-rata gini rasio berpengaruh secara positif terhadap tingkat pengangguran terbuka di kabupaten/kota wilayah Provinsi Sumatera Utara, yang cenderung berada di wilayah utara (atas) dan selatan (bawah) pulau Sumatera Utara. Sedangkan, warna merah muda (plot kanan) menandakan rata-rata gini rasio berpengaruh secara negatif terhadap tingkat pengangguran terbuka di kabupaten/kota wilayah Provinsi Sumatera Utara, yang cenderung berada di wilayah tengah pulau Sumatera Utara.

4.7.5. Keragaman Spasial Taksiran β_4

Gambar 4.11 menunjukkan sebaran nilai taksiran β_4 , beserta $p - value$ -nya, yang menggambarkan pengaruh rata-rata laju pertumbuhan penduduk (X_4) terhadap tingkat pengangguran terbuka (Y) di setiap kabupaten/kota di Sumatera Utara.

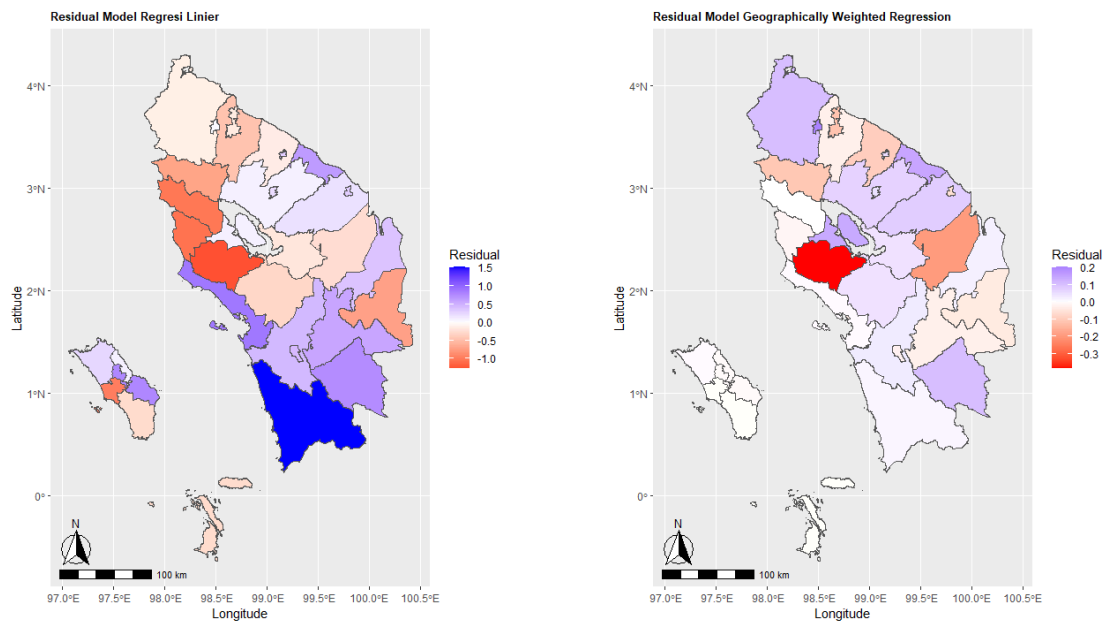


Gambar 4.11 Peta Keragaman Spasial Taksiran β_4

Di beberapa kabupaten/kota yang saling berdekatan, dapat dilihat bahwa nilai taksiran parameternya hampir sama. Warna biru muda (plot kanan) menandakan rata-rata laju pertumbuhan penduduk berpengaruh secara positif terhadap tingkat pengangguran terbuka di kabupaten/kota wilayah Provinsi Sumatera Utara, yang cenderung berada di wilayah selatan (bawah) pulau Sumatera Utara. Sedangkan, warna merah muda (plot kiri) menandakan rata-rata laju pertumbuhan penduduk berpengaruh secara negatif terhadap tingkat pengangguran terbuka di kabupaten/kota wilayah Provinsi Sumatera Utara, yang cenderung berada di wilayah selatan (bawah) pulau Sumatera Utara. Secara umum, variabel laju pertumbuhan penduduk berpengaruh signifikan terhadap tingkat pengangguran terbuka di Sumatera Utara tahun 2022, ditunjukkan dengan banyak wilayah kabupaten/kota yang berwarna mendekati putih (plot kiri).

4.7.6. Perbandingan Residual Regresi Global dan Regresi Lokal

Gambar 4.12 menunjukkan perbandingan residual dari model regresi global (regresi linier) dan regresi lokal (GWR) dalam memodelkan tingkat pengangguran terbuka di Sumatera Utara tahun 2022.



Gambar 4.12 Grafik Perbandingan Residual Regresi Global dan Regresi Lokal

Terlihat bahwa regresi lokal lebih baik dalam memodelkan tingkat pengangguran terbuka di Sumatera Utara tahun 2022 karena pada plotnya terlihat warna dari residualnya relatif banyak yang pudar. Hal tersebut menandakan bahwa banyak residual yang mendekati 0.

BAB IV

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Terjadi heterogenitas spasial pada data tingkat pengangguran terbuka di Sumatera Utara tahun 2022.
2. Dari hasil penelitian, didapati model terbaik untuk memodelkan kasus tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022 adalah model *Geographically Weighted Regression* (GWR) Kernel Tricube dengan *adaptive bandwidth*.
3. Faktor-faktor yang berpengaruh pada tingkat pengangguran terbuka di Sumatera Utara pada tahun 2022 cenderung berbeda-beda pada tiap kabupaten/kota dan besarnya cenderung membentuk kelompok pada suatu wilayah.

DAFTAR PUSTAKA

- [1] Adi Putra, I. K. A. (n.d.). Analisis Pengaruh Tingkat Pengangguran Terbuka, Kesempatan Kerja, Dan Tingkat Pendidikan Terhadap Tingkat Kemiskinan Pada Kabupaten/Kota Di Provinsi Bali. *E-Jurnal EP Unud*, 7(3), 416-444.
- [2] BPS Sumatera Utara. (2023). *Gini Ratio Sumatera Utara Menurut Kabupaten/Kota 2021-2023*.
<https://sumut.bps.go.id/indicator/23/467/1/gini-ratio-sumatera-utara-menurut-kabupaten-kota.html>
- [3] BPS Sumatera Utara. (2023). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk, Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota, 2020 dan 2022*.
<https://sumut.bps.go.id/statictable/2023/03/10/2927/penduduk-laju-pertumbuhan-penduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelamin-penduduk-menurut-kabupaten-kota-2020-dan-2022.html>
- [4] BPS Sumatera Utara. (2023). *Tingkat Pengangguran Terbuka (TPT) Penduduk Umur 15 Tahun Keatas Menurut Kab/Kota (Persen), 2021-2023*.
<https://sumut.bps.go.id/indicator/6/44/1/tingkat-pengangguran-terbuka-tpt-penduduk-umur-15-tahun-keatas-manurut-kab-kota.html>
- [5] Ghozali, I. (2016). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 23* (Edisi 8). Cetakan ke VIII. Badan Penerbit Universitas Diponegoro.
- [6] Fotheringham, A. S., et al. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- [7] Open Data Sumatera Utara. (2022). *Rasio Ketergantungan (Dependency Ratio) Menurut Kabupaten/Kota Di Provinsi Sumatera Utara, 2022*.
<https://sadaina.sumutprov.go.id/open-data/dataset/2d92576a-126d-51c1-a8bb-5a8bd9c7b8e8>
- [8] Safitri, U., & Amaliana, L. (2021). Model Geographically Weighted Regression Dengan Fungsi Pembobot Adaptive dan Fixed Kernel Pada Kasus Kematian Ibu Di Jawa Timur. *Jurnal Statistika Dan Aplikasinya*, 5, no. 2.
- [9] Suryono. (2015). *Analisis Regresi untuk Penelitian* (1st ed. ed., Vol. Vol.1). Deepublish.

- [10] Uthami, I. A. (2013). Regresi Kuantil Median untuk Mengatasi Heteroskedastisitas pada Analisis Regresi. *e-Jurnal Matematika*, 2(1), 6-13.
- [11] Yulianti, S. (2020). *Pemodelan Data Panel Spasial Pada Indeks Pembangunan Manusia Di Jawa Tengah*. Universitas Indonesia Library, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

LAMPIRAN *CODE*

〇〇〇

```
# Data untuk membuat peta
library(sf)
# https://geosai.my.id/download-shp-kabupaten-kota-indonesia/
maps <- st_read("D:/Materi Kuliah UI/Topik Khusus I - Analisis Data Spasial/Tugas Topik Khusus I - Analisis Data
Spasial/Sumatera_Utara_ADMIN_BPS.shp")
maps$Kabupaten <- gsub("Kota ", "", maps$Kabupaten)
maps <- maps[order(maps$Kabupaten), ]
class(maps)
library(ggplot2)
ggplot(maps) + geom_sf()

# Data Tingkat Pengangguran Terbuka
library(readxl)
df <- read_excel("D:/Materi Kuliah UI/Topik Khusus I - Analisis Data Spasial/Tugas Topik Khusus I - Analisis Data
Spasial/data_sumut.xlsx")
df <- data.frame(df)
colnames(df)[1] <- "kab_kota"
df <- df[-1, ] # Menghilangkan Sumatera Utara
df <- df[order(df$kab_kota), ]
df[, -1] <- lapply(df[, -1], as.numeric)
View(df)

# Preprocessing
# Periksa nama kabupaten/kota
maps <- maps[!maps$Kabupaten=="Danau Toba", ]
df$kab_kota==maps$Kabupaten
df[df$kab_kota=="Labuhan Batu", ]$kab_kota <- "Labuhanbatu"
df[df$kab_kota=="Labuhanbatu Utara", ]$kab_kota <- "Labuhanbatu Utara"
maps[maps$Kabupaten=="Labuhan Batu", ]$Kabupaten <- "Labuhanbatu"
maps[maps$Kabupaten=="Labuhan Batu Selatan", ]$Kabupaten <- "Labuhanbatu Selatan"
maps[maps$Kabupaten=="Labuhan Batu Utara", ]$Kabupaten <- "Labuhanbatu Utara"
maps[maps$Kabupaten=="Tanjung Balai", ]$Kabupaten <- "Tanjungbalai"
maps[maps$Kabupaten=="Pematang Siantar", ]$Kabupaten <- "Pematangsiantar"
maps[maps$Kabupaten=="Toba Samosir", ]$Kabupaten <- "Toba"
df <- df[order(df$kab_kota), ]
maps <- maps[order(maps$Kabupaten), ]
colnames(maps)[5] <- "kab_kota"
df$kab_kota==maps$kab_kota

# Merge data
library(dplyr)
df_merge_sp <- inner_join(maps, df, by = "kab_kota")
df_merge_sp <- as(df_merge_sp, "Spatial")

# Cari centroid koordinat
centroid <- st_centroid(maps)
coords <- st_coordinates(centroid)
# Cek centroid pada peta
ggplot(maps) + geom_sf(aes(fill = df$kab_kota), show.legend = FALSE) +
  geom_sf(data = centroid, col = "black")
# Save koordinat
df_sp <- SpatialPointsDataFrame(coords, df)

# Label plot
label <- c()
i <- 1
for(lab in df$kab_kota){
  if (length(strsplit(lab, " ")[[1]]) > 1 || nchar(lab)>4){
    # Extract first three-letter string
    first_three <- gsub("\\b(\\w{3})\\w*\\b", "\\1", lab)
    first_three <- gsub("\\s", "", first_three)
    label[i] <- first_three
  } else{
    label[i] <- df$kab_kota[i]
  }
  i <- i + 1
}
label
labels <- data.frame(label = label, long = coords[, 1], lat = coords[, 2])

# Statistika Deskriptif
summary(df)
df[df$TPT_2022==min(df$TPT_2022), ]
df[df$TPT_2022==max(df$TPT_2022), ]
```

○○○

```
# Tingkat Pengangguran Terbuka (TPT) Tahun 2022
library(ggspatial)
df[order(df$TPT_2022), ]
p1 <- ggplot(maps) + geom_sf(aes(fill = df$TPT_2022)) +
  geom_text(data = labels, aes(x = long, y = lat, label = label), size = 4, color = "black") +
  scale_fill_gradient(name = "TPT", low = "lightblue", high = "darkblue") +
  ggtitle("Tingkat Pengangguran Terbuka (TPT) \ndi Provinsi Sumatera Utara Tahun 2022") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
# https://bookdown.org/brianwood1/QDA55/simple-static-maps.html
p1

# Persentase Penduduk Miskin (PPM) Tahun 2022
p2 <- ggplot(maps) + geom_sf(aes(fill = df$PPM_2022)) +
  scale_fill_gradient(name = "PPM", low = "lightgreen", high = "darkgreen") +
  ggtitle("Persentase Penduduk Miskin (PPM) \ndi Provinsi Sumatera Utara Tahun 2022") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
p2

# Dependency Ratio (DR) Tahun 2022
p3 <- ggplot(maps) + geom_sf(aes(fill = df$DR_2022)) +
  scale_fill_gradient(name = "DR", low = "yellow", high = "red") +
  ggtitle("Dependency Ratio (DR) \ndi Provinsi Sumatera Utara Tahun 2022") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
p3

# Gini Rasio (GR) Tahun 2022
library(viridis)
p4 <- ggplot(maps) + geom_sf(aes(fill = df$GR_2022)) +
  scale_fill_viridis_c(option = "cividis", name = "GR", direction = -1) +
  ggtitle("Gini Rasio (GR) \ndi Provinsi Sumatera Utara Tahun 2022") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
p4

# Laju Pertumbuhan Penduduk (LPP) Tahun 2022
p5 <- ggplot(maps) + geom_sf(aes(fill = df$LPP_2022)) +
  scale_fill_viridis_c(option = "rocket", name = "LPP", direction = -1) +
  ggtitle("Laju Pertumbuhan Penduduk (LPP) \ndi Provinsi Sumatera Utara Tahun 2022") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
p5

# Plot Transformasi Variabel
df$log_TPT <- log(df$TPT_2022)
density_data1 <- data.frame(x = df$TPT_2022, group = "Sebelum")
density_data2 <- data.frame(x = log(df$TPT_2022), group = "Sesudah")
# Combine the datasets
density_data <- rbind(density_data1, density_data2)
# Create the plot
ggplot(density_data, aes(x = x, fill = group)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("blue", "red")) +
  labs(title = "Perbandingan Desnitas Sebelum dan Sesudah \nTransformasi Logaritma Natural",
    x = "Values",
    y = "Density") +
  theme_minimal()

# Regresi Linier
reg <- lm(log(TPT_2022) ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022, data = df)
summary(reg)
AIC(reg)
```

○○○

```
# Uji Asumsi
# Uji Multikolinearitas
library(car)
vif(reg)
cor(df[, -1])

# Uji Normalitas
qqnorm(reg$residuals, ylab = "Residuals", xlab = "Normal Scores")
qqline(reg$residuals)
shapiro.test(reg$residuals)
ks.test(reg$residuals, "pnorm")

# Uji Durbin Watson (Autocorrelation)
dwtest(log(TPT_2022) ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022, data = df)

# Uji Heteroskedastisitas
library(lmtest)
bptest(reg)
spreadLevelPlot(reg)
par(mfrow = c(2, 2))
plot(reg)
par(mfrow = c(1, 1))
plot(reg, 1)

# Geographically Weighted Regression (GWR)
# https://crd230.github.io/gwr.html
library(spgwr)

# Estimated Optimal Bandwidth
# Gaussian Fixed Bandwidth
gwr_b1 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = FALSE)

gwr_b1
# Fit Model
gwr_fit1 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b1, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit1

# Gaussian Adaptive Bandwidth
gwr_b2 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = TRUE)

gwr_b2
# Fit Model
gwr_fit2 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b2, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit2

# Bisquare Fixed Bandwidth
gwr_b3 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = FALSE, gweight = gwr.bisquare)

gwr_b3
# Fit Model
gwr_fit3 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b3, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit3

# Bisquare Adaptive Bandwidth
gwr_b4 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = TRUE, gweight = gwr.bisquare)

gwr_b4
# Fit Model
gwr_fit4 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b4, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit4

# Tricube Fixed Bandwidth
gwr_b5 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = FALSE, gweight = gwr.tricube)

gwr_b5
# Fit Model
gwr_fit5 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b5, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit5

# Tricube Adaptive Bandwidth
gwr_b6 <- gwr.sel(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
                 data = df, coords = coords, adapt = TRUE, gweight = gwr.tricube)

gwr_b6
# Fit Model
gwr_fit6 <- gwr(log_TPT ~ PPM_2022 + DR_2022 + GR_2022 + LPP_2022,
               data = df, coords = coords, bandwidth = gwr_b6, se.fit = TRUE, hatmatrix = TRUE)

gwr_fit6
```

〇〇〇

```
# Model Terbaik
# AIC
AIC(reg)
gwr_fit1$results$AIC
gwr_fit2$results$AIC
gwr_fit3$results$AIC
gwr_fit4$results$AIC
gwr_fit5$results$AIC
gwr_fit6$results$AIC

# Untuk model pilihan:
gwr_fit <- gwr_fit6
gwr_fit

# Statistical Test
dfree <- gwr_fit$results$edf

df$PPM_2022_t <- gwr_fit$SDF$PPM_2022/gwr_fit$SDF$PPM_2022_se
df$PPM_2022_t.pval <- 2*pt(-abs(df$PPM_2022_t), dfree)
df$PPM_2022_t.signif <- ifelse(df$PPM_2022_t.pval<0.05, "Signifikan", "Tidak Signifikan")

df$DR_2022_t <- gwr_fit$SDF$DR_2022/gwr_fit$SDF$DR_2022_se
df$DR_2022_t.pval <- 2*pt(-abs(df$DR_2022_t), dfree)
df$DR_2022_t.signif <- ifelse(df$DR_2022_t.pval<0.05, "Signifikan", "Tidak Signifikan")

df$GR_2022_t <- gwr_fit$SDF$GR_2022/gwr_fit$SDF$GR_2022_se
df$GR_2022_t.pval <- 2*pt(-abs(df$GR_2022_t), dfree)
df$GR_2022_t.signif <- ifelse(df$GR_2022_t.pval<0.05, "Signifikan", "Tidak Signifikan")

df$LPP_2022_t <- gwr_fit$SDF$LPP_2022/gwr_fit$SDF$LPP_2022_se
df$LPP_2022_t.pval <- 2*pt(-abs(df$LPP_2022_t), dfree)
df$LPP_2022_t.signif <- ifelse(df$LPP_2022_t.pval<0.05, "Signifikan", "Tidak Signifikan")

# Kelompok variabel signifikan
x1 <- df[df$PPM_2022_t.signif=="Signifikan", ]$kab_kota
length(x1)
x1
x2 <- df[df$DR_2022_t.signif=="Signifikan", ]$kab_kota
length(x2)
x2
x3 <- df[df$GR_2022_t.signif=="Signifikan", ]$kab_kota
length(x3)
x3
x4 <- df[df$LPP_2022_t.signif=="Signifikan", ]$kab_kota
length(x4)
x4
df_sig <- data.frame(kab_kota = df$kab_kota)
rownames(df_sig) <- df_sig$kab_kota
df_sig$x1 <- ifelse(df_sig$kab_kota %in% x1, "x1", "0")
df_sig$x2 <- ifelse(df_sig$kab_kota %in% x2, "x2", "0")
df_sig$x3 <- ifelse(df_sig$kab_kota %in% x3, "x3", "0")
df_sig$x4 <- ifelse(df_sig$kab_kota %in% x4, "x4", "0")
df_sig <- df_sig[, -1]
head(df_sig)
tabulasi <- apply(df_sig, 1, function(x) paste(x, collapse = "_"))
df_sig$tabulasi <- tabulasi
table(df_sig$tabulasi)
list_nama <- split(rownames(df_sig), df_sig$tabulasi)
length(list_nama)
nama <- c(1, 4, 3, 6, 2, 5, 8, 7, 9)
names(nama) <- names(list_nama)
library(plyr)
df_sig$kelompok <- plyr::mapvalues(df_sig$tabulasi, from = names(nama), to = nama)
View(df_sig)

# Visualisasi
View(maps)
ggplot(maps) + geom_sf(aes(fill = df_sig$kelompok)) +
  geom_text(data = labels, aes(x = long, y = lat, label = label), size = 4, color = "black") +
  ggtitle("Kelompok Kabupaten/Kota Berdasarkan \nVariabel yang Signifikan") +
  xlab("Longitude") + ylab("Latitude") + labs(fill = "Kelompok") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 14, face = "bold"))
```

○ ○ ○

```
# Visualisasi P-value Koefisien
breaks <- c(0, 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1)
# PPM
ppm_class <- cut(df$PPM_2022_t.pval, breaks)
unique(ppm_class)
ppm_class
k11 <- ggplot(maps) + geom_sf(aes(fill = df$PPM_2022_t.pval)) +
  scale_fill_gradient(name = "p-value", low = "white", high = "blue",
    breaks = breaks, labels = breaks) +
  ggtitle("P-value Koefisien Persentase Penduduk Miskin (PPM)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# DR
k12 <- ggplot(maps) + geom_sf(aes(fill = df$DR_2022_t.pval)) +
  scale_fill_gradient(name = "p-value", low = "white", high = "blue",
    breaks = breaks, labels = breaks) +
  ggtitle("P-value Koefisien Dependency Ratio (DR)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# GR
k13 <- ggplot(maps) + geom_sf(aes(fill = df$GR_2022_t.pval)) +
  scale_fill_gradient(name = "p-value", low = "white", high = "blue",
    breaks = breaks, labels = breaks) +
  ggtitle("P-value Koefisien Gini Rasio (GR)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# LPP
k14 <- ggplot(maps) + geom_sf(aes(fill = df$LPP_2022_t.pval)) +
  scale_fill_gradient(name = "p-value", low = "white", high = "blue",
    breaks = breaks, labels = breaks) +
  ggtitle("P-value Koefisien Laju Pertambahan Penduduk (LPP)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))

# Visualisasi Koefisien
# PPM
k21 <- ggplot(maps) + geom_sf(aes(fill = gwr_fit$SDF$PPM_2022)) +
  scale_fill_gradient2(name = "Koefisien", low = "red", mid = "white", high = "blue") +
  ggtitle("Koefisien Persentase Penduduk Miskin (PPM)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# DR
k22 <- ggplot(maps) + geom_sf(aes(fill = gwr_fit$SDF$DR_2022)) +
  scale_fill_gradient2(name = "Koefisien", low = "red", mid = "white", high = "blue") +
  ggtitle("Koefisien Dependency Ratio (DR)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# GR
k23 <- ggplot(maps) + geom_sf(aes(fill = gwr_fit$SDF$GR_2022)) +
  scale_fill_gradient2(name = "Koefisien", low = "red", mid = "white", high = "blue") +
  ggtitle("Koefisien Gini Rasio (GR)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# LPP
k24 <- ggplot(maps) + geom_sf(aes(fill = gwr_fit$SDF$LPP_2022)) +
  scale_fill_gradient2(name = "Koefisien", low = "red", mid = "white", high = "blue") +
  ggtitle("Koefisien Laju Pertambahan Penduduk (LPP)") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
```

○ ○ ○

```
# Gabungan Visualisasi
library(ggpubr)
ggarrange(k11, k21, ncol = 2)
ggarrange(k12, k22, ncol = 2)
ggarrange(k13, k23, ncol = 2)
ggarrange(k14, k24, ncol = 2)

# Visualisasi Residual
# Regresi Linier
r1 <- ggplot(maps) + geom_sf(aes(fill = reg$residuals)) +
  scale_fill_gradient2(name = "Residual", low = "red", mid = "white",
    high = "blue", midpoint = 0) +
  ggtitle("Residual Model Regresi Linier") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
# GWR
r2 <- ggplot(maps) + geom_sf(aes(fill = gwr_fit$SDF$gwr.e)) +
  scale_fill_gradient2(name = "Residual", low = "red", mid = "white",
    high = "blue", midpoint = 0) +
  ggtitle("Residual Model Geographically Weighted Regression") +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.4) +
  annotation_north_arrow(location = "bl", which_north = "true",
    pad_x = unit(0.0, "in"), pad_y = unit(0.2, "in"),
    style = north_arrow_fancy_orienteering) +
  theme(plot.title = element_text(size = 10, face = "bold"))
ggarrange(r1, r2, ncol = 2)
```


LAMPIRAN DATA

kab_kota	TPT_2022	PPM_2022	DR_2022	GR_2022	LPP_2022	longitude	latitude
Asahan	6.26	8.64	42.75	0.255	1.31	99.56282855	2.802542074
Batu Bara	6.21	11.53	46.85	0.243	0.79	99.47563895	3.234237548
Binjai	6.36	5.1	42.32	0.302	1.59	98.48871083	3.599491152
Dairi	0.88	7.88	48.16	0.221	1.23	98.26731654	2.84362578
Deli Serdang	8.79	3.62	42.13	0.27	0.67	98.69435114	3.480842731
Gunungsitoli	3.65	14.81	46.19	0.303	0.66	97.58309154	1.265094099
Humbang Hasundutan	0.42	8.86	50.82	0.258	1.31	98.5724341	2.270728046
Karo	2.71	8.17	42.43	0.236	1.32	98.29604788	3.120686547
Labuhanbatu	6.9	8.26	42.2	0.278	1.62	100.0564698	2.279357048
Labuhanbatu Selatan	3.15	8.09	42.4	0.247	1.13	100.1305356	1.83886421
Labuhanbatu Utara	3.75	9.09	43.14	0.276	1.33	99.7546035	2.409090453
Langkat	6.88	9.49	43.65	0.26	0.54	98.22034671	3.710688448
Mandailing Natal	7.64	8.92	49.27	0.244	1.44	99.36944505	0.77436529
Medan	8.89	8.07	39.51	0.399	1.38	98.66799658	3.627720856
Nias	2.81	16	49.93	0.275	1	97.72837854	1.065937681
Nias Barat	0.53	24.75	45.07	0.21	0.86	97.4896517	1.025365843
Nias Selatan	3.69	16.48	35.18	0.223	2.07	98.10363891	0.32041484
Nias Utara	2.59	23.4	41.27	0.239	1.35	97.35547868	1.312779528
Padang Lawas	5.9	8.05	46.57	0.227	1.36	99.84412445	1.140399419
Padang Lawas Utara	4.31	8.94	46.9	0.245	1.43	99.75745406	1.59325322
Padangsidempuan	7.76	6.89	43.06	0.275	1.5	99.28079543	1.396311455
Pakpak Bharat	0.26	8.66	49.85	0.274	2.44	98.24060676	2.55417163
Pematangsiantar	9.36	7.88	41.68	0.321	1.23	99.0589741	2.962597604
Samosir	1.16	11.77	52.46	0.298	1.21	98.72129157	2.558723482
Serdang Bedagai	4.98	7.82	44.51	0.24	0.91	99.05853505	3.367155536
Sibolga	7.05	11.47	42.49	0.346	0.5	98.78590557	1.738364364
Simahungun	5.51	8.26	40.62	0.262	1.8	99.03322489	2.977031356
Tanjungbalai	4.62	12.45	44.18	0.245	1.2	99.79498813	2.964415352
Tapanuli Selatan	3.65	8.07	48.33	0.209	1.21	99.25804545	1.492726215
Tapanuli Tengah	7.97	11.71	43.13	0.233	1.49	98.65979448	1.833887203
Tapanuli Utara	1.07	8.93	52.18	0.242	1.03	99.06974134	1.989761642
Tebing Tinggi	6.39	9.59	42.08	0.334	1.63	99.15570836	3.328360051
Toba	1.39	8.89	48.93	0.305	1.63	99.2642168	2.379730791

KONTRIBUSI KELOMPOK

No.	Nama	NPM	Kontribusi	Tingkat Kontribusi
1.	Daniella Putri Shalomita	2106631072	Menyusun makalah dan poster	100%
2.	Diki Wahyudi	2106709131	Melakukan pengolahan dan analisis data	100%
3.	Raqi Akbar Robbani	2106652562	Menyusun makalah dan poster	100%