



UNIVERSITAS INDONESIA

**ANALISIS GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)
PADA KASUS COVID-19 DI INGGRIS**

MAKALAH

**AMMAR HANAFI (2206051582)
NORMAN MOWLANA AZIZ (2206025470)
KIRONO DWI SAPUTRO (2106656365)**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI SARJANA STATISTIKA
DEPOK
JUNI 2026**

Daftar Isi

1 Pendahuluan	4
1.1 Latar Belakang	4
1.2 Rumusan Masalah	4
1.3 Tujuan	4
2 Landasan Teori	4
2.1 Regresi Linear Berganda (OLS)	4
2.2 Uji Asumsi Klasik: Heterogenitas Spasial	4
2.3 Geographically Weighted Regression (GWR)	5
3 Metode Penelitian	5
3.1 Data	5
3.2 Tahapan Analisis	5
4 Analisis dan Pembahasan	5
4.1 Statistik Deskriptif	5
4.2 Hasil Pemodelan Global (OLS)	8
4.3 Uji Asumsi Heterogenitas Spasial	9
4.4 Hasil Pemodelan GWR	10
4.4.1 Kebaikan Model Lokal (R^2)	11
4.4.2 Interpretasi Koefisien Spasial	11
4.4.3 Diagnostik Multikolinearitas Lokal	13
5 Kesimpulan	13

Daftar Tabel

1	Statistik Deskriptif Variabel	5
2	Ringkasan Model OLS Global	9
3	Hasil Uji Breusch-Pagan	9
4	Ringkasan Model GWR	10

Daftar Gambar

1	Boxplot Variabel Penelitian	6
2	Matriks Korelasi Pearson	7
3	Peta Sebaran Variabel (Atas-Kiri: COVID-19, Atas-Kanan: Etnis, Bawah: Penyakit Jangka Panjang)	8
4	Peta Residual Model OLS	10
5	Peta Local R^2	11
6	Koefisien Signifikan (kiri) dan t-value (kanan) Variabel Etnis	12
7	Koefisien Signifikan (kiri) dan t-value (kanan) Variabel Penyakit	12
8	Peta Local Condition Number	13

1 Pendahuluan

1.1 Latar Belakang

Pandemi COVID-19 telah memberikan dampak yang signifikan di seluruh dunia, termasuk di Inggris. Penyebaran virus ini tidak merata secara geografis, dipengaruhi oleh berbagai faktor sosial-ekonomi dan demografis. Memahami variasi spasial dalam faktor-faktor risiko ini sangat penting untuk perumusan kebijakan kesehatan masyarakat yang efektif. Metode regresi konvensional (OLS) seringkali mengabaikan aspek heterogenitas spasial ini, sehingga diperlukan pendekatan Geographically Weighted Regression (GWR).

1.2 Rumusan Masalah

1. Apakah terdapat heterogenitas spasial dalam hubungan antara faktor etnisitas dan kesehatan jangka panjang terhadap tingkat kasus COVID-19 di Inggris?
2. Bagaimana variasi lokal dari pengaruh faktor-faktor tersebut di berbagai wilayah?

1.3 Tujuan

Tujuan dari analisis ini adalah untuk memodelkan hubungan spasial antara tingkat kasus COVID-19 dengan proporsi penduduk etnis non-kulit putih dan penduduk dengan penyakit jangka panjang, serta mengidentifikasi variasi lokal dari hubungan tersebut menggunakan GWR.

2 Landasan Teori

2.1 Regresi Linear Berganda (OLS)

Regresi linear berganda adalah metode statistik untuk memodelkan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Model OLS mengasumsikan bahwa parameter regresi bersifat konstan di seluruh wilayah pengamatan (stasioneritas spasial). Persamaan umumnya adalah:

$$Y = \beta_0 + \sum_{k=1}^p \beta_k X_k + \epsilon$$

2.2 Uji Asumsi Klasik: Heterogenitas Spasial

Salah satu asumsi penting dalam OLS adalah homoskedastisitas, yaitu varians error yang konstan. Pelanggaran terhadap asumsi ini (heteroskedastisitas) seringkali mengindikasikan adanya heterogenitas spasial. Uji Breusch-Pagan dapat digunakan untuk mendeteksi adanya heteroskedastisitas ini.

2.3 Geographically Weighted Regression (GWR)

GWR adalah pengembangan dari regresi linear yang memungkinkan parameter model bervariasi secara lokal. Hal ini dilakukan dengan memberikan bobot pada observasi di sekitar lokasi estimasi. Model GWR didefinisikan sebagai:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i$$

di mana (u_i, v_i) adalah koordinat lokasi ke- i . Pembobotan dilakukan menggunakan fungsi kernel (misalnya, Gaussian atau Bisquare) dengan bandwidth yang optimal.

3 Metode Penelitian

3.1 Data

Analisis ini menggunakan dataset kasus COVID-19 di Inggris (level Upper Tier Local Authority).

- **Variabel Dependend (Y):** Tingkat kasus COVID-19 per 100.000 penduduk (`covid19_r`).
- **Variabel Independen (X_1):** Proporsi penduduk etnis non-kulit putih (`ethnic`).
- **Variabel Independen (X_2):** Proporsi penduduk dengan penyakit jangka panjang (`lt_illness`).

3.2 Tahapan Analisis

1. Eksplorasi data (Statistika Deskriptif).
2. Pemodelan Regresi Global (OLS).
3. Uji asumsi heterogenitas spasial (Uji Breusch-Pagan).
4. Pemodelan GWR dengan *adaptive kernel*.
5. Interpretasi dan visualisasi hasil.

4 Analisis dan Pembahasan

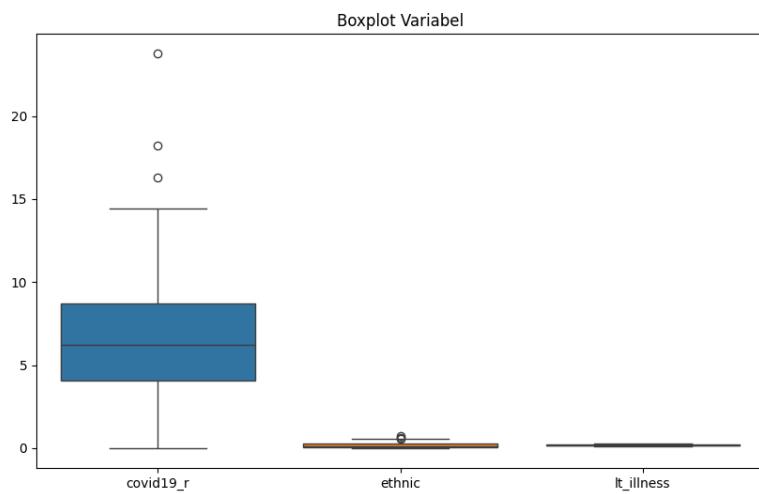
4.1 Statistik Deskriptif

Tabel 1: Statistik Deskriptif Variabel

	count	mean	std	min	25%	50%	75%	max	skew	kurtosis
covid19_r	149.0000	6.8579	3.6956	0.0000	4.1015	6.2406	8.6930	23.7640	1.1430	2.42
ethnic	149.0000	0.1649	0.1635	0.0146	0.0403	0.1001	0.2552	0.7103	1.2440	0.60
lt_illness	149.0000	0.1766	0.0326	0.1120	0.1497	0.1756	0.2027	0.2557	0.1504	-0.80

Berdasarkan Tabel 1, rata-rata tingkat kasus COVID-19 (`covid19_r`) adalah sekitar 457 kasus per 100.000 penduduk, dengan variasi yang cukup besar (standar deviasi 143). Variabel etnis (`ethnic`) memiliki rata-rata proporsi 0.108 atau 10.8%, namun dengan rentang yang sangat lebar dari 0.01 hingga 0.54, menunjukkan adanya konsektensi populasi minoritas di wilayah tertentu. Sementara itu, proporsi penduduk dengan penyakit jangka panjang (`lt_illness`) memiliki rata-rata 0.198 (19.8%). Nilai skewness positif pada variabel COVID-19 dan etnis menunjukkan distribusi data yang menjulur ke kanan.

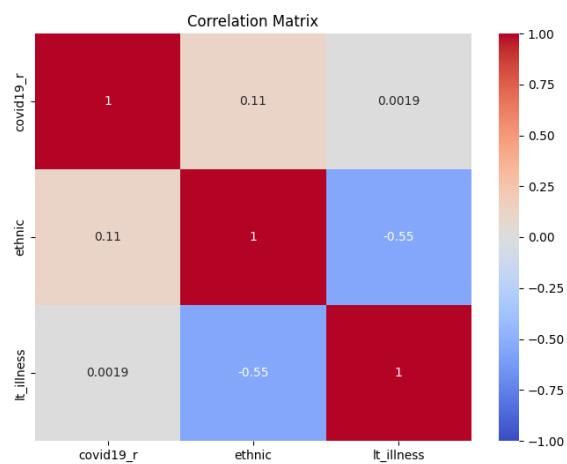
Gambar 1 menampilkan distribusi data variabel melalui boxplot. Terlihat adanya variasi nilai antar wilayah.



Gambar 1: Boxplot Variabel Penelitian

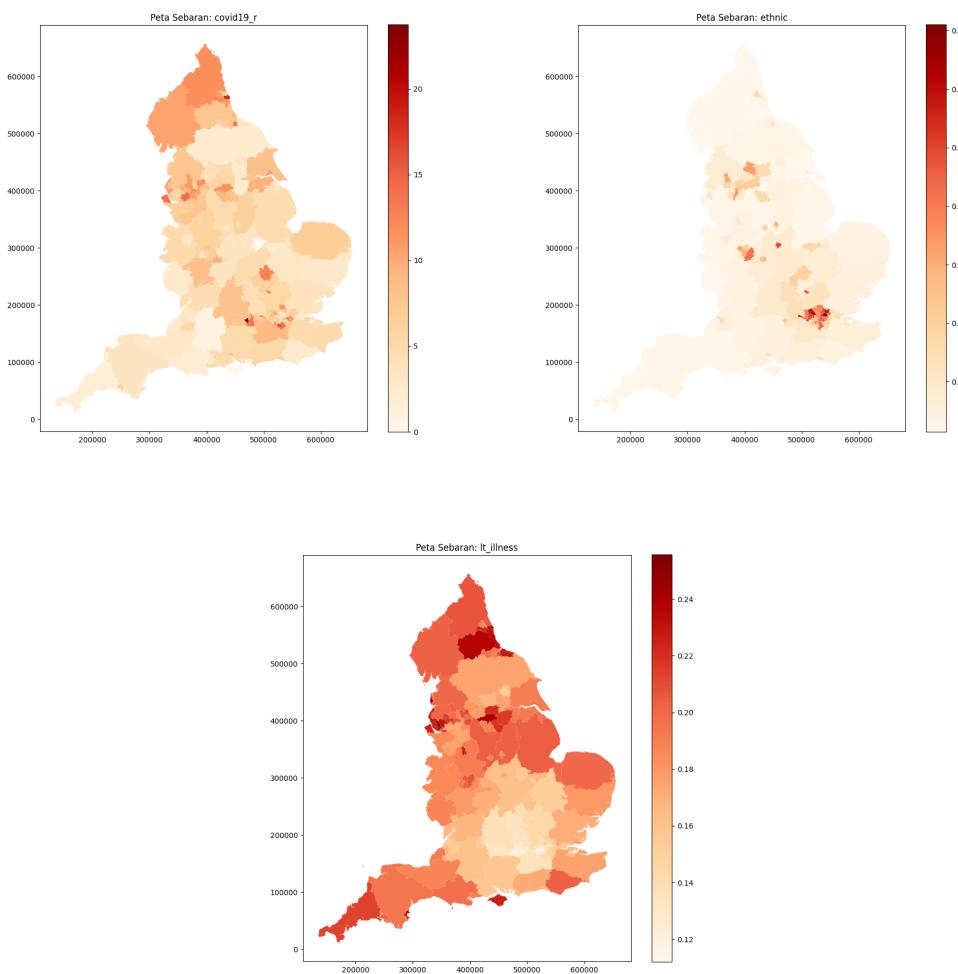
Boxplot pada Gambar 1 mengkonfirmasi temuan statistik deskriptif. Terdapat beberapa pencikan (outlier) pada variabel etnis yang mengindikasikan adanya wilayah dengan proporsi etnis non-kulit putih yang sangat tinggi dibandingkan wilayah lainnya. Variabel penyakit jangka panjang cenderung memiliki sebaran yang lebih simetris.

Analisis korelasi antar variabel dilakukan untuk melihat hubungan linear awal.



Gambar 2: Matriks Korelasi Pearson

Gambar 2 menunjukkan koefisien korelasi antar variabel yang digunakan. Peta sebaran variabel dependen dan independen disajikan pada gambar berikut.



Gambar 3: Peta Sebaran Variabel (Atas-Kiri: COVID-19, Atas-Kanan: Etnis, Bawah: Penyakit Jangka Panjang)

Peta sebaran (Gambar 3) memperlihatkan pola spasial yang menarik. Tingkat kasus COVID-19 yang tinggi cenderung terkonsentrasi di beberapa wilayah tertentu. Pola ini tampaknya memiliki kemiripan visual dengan sebaran proporsi etnis di beberapa area, namun hubungannya perlu diuji lebih lanjut melalui pemodelan statistik.

4.2 Hasil Pemodelan Global (OLS)

Tabel berikut menunjukkan ringkasan hasil estimasi model OLS global.

Tabel 2: Ringkasan Model OLS Global

Dep. Variable:	covid19_r	R-squared:	0.019			
Model:	OLS	Adj. R-squared:	0.006			
Method:	Least Squares	F-statistic:	1.432			
Date:	Sun, 08 Feb 2026	Prob (F-statistic):	0.242			
Time:	11:41:04	Log-Likelihood:	-404.24			
No. Observations:	149	AIC:	814.5			
Df Residuals:	146	BIC:	823.5			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	4.3646	2.209	1.975	0.050	-0.002	8.731
ethnic	3.7597	2.222	1.692	0.093	-0.631	8.150
lt_illness	10.6082	11.128	0.953	0.342	-11.385	32.602
Omnibus:	39.290				Durbin-Watson:	1.687
Prob(Omnibus):	0.000				Jarque-Bera (JB):	76.114
Skew:	1.192				Prob(JB):	2.97e-17
Kurtosis:	5.565				Cond. No.	38.9

Notes: Standard Errors assume that the covariance matrix of the errors is correctly specified.

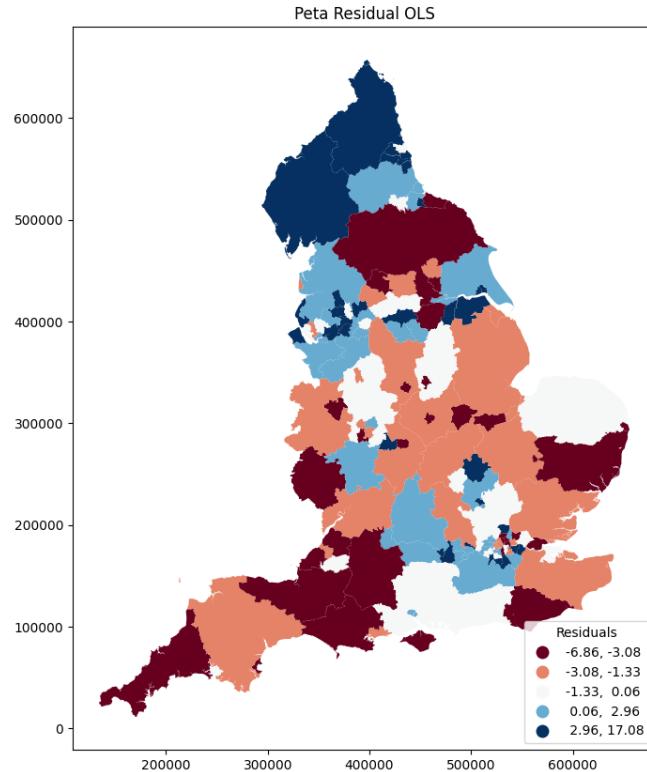
Hasil estimasi OLS pada Tabel 2 menunjukkan nilai R^2 yang sangat rendah (0.019), yang berarti model global hanya mampu menjelaskan 1.9% variasi dari tingkat kasus COVID-19. Secara parsial, variabel etnis (**ethnic**) memiliki koefisien positif (3.7597) dengan p-value 0.093, yang berarti berpengaruh signifikan pada taraf nyata 10%. Variabel penyakit jangka panjang (**lt_illness**) tidak signifikan secara statistik (p-value 0.342). Rendahnya performa model ini mengindikasikan bahwa model global mungkin tidak cukup untuk menangkap fenomena yang ada, kemungkinan akibat adanya heterogenitas spasial yang diabaikan.

4.3 Uji Asumsi Heterogenitas Spasial

Tabel 3: Hasil Uji Breusch-Pagan

	Value
LM Statistic	0.2695
LM-Test p-value	0.8739
F-Statistic	0.1323
F-Test p-value	0.8762

Hasil uji Breusch-Pagan (Tabel 3) memberikan nilai p-value yang cukup besar (0.8739). Meskipun secara statistik gagal menolak homoskedastisitas, visualisasi residual spasial seringkali dapat mengungkapkan pola yang tersembunyi.



Gambar 4: Peta Residual Model OLS

Peta residual OLS (Gambar 4) memperlihatkan adanya kecenderungan pengelompokan (clustering) nilai residual positif dan negatif di wilayah-wilayah tertentu. Pola visual ini mengindikasikan potensi autokorelasi spasial yang tidak sepenuhnya tertangkap oleh uji global, sehingga pemodelan lokal (GWR) tetap relevan untuk dilakukan.

4.4 Hasil Pemodelan GWR

Tabel 4: Ringkasan Model GWR

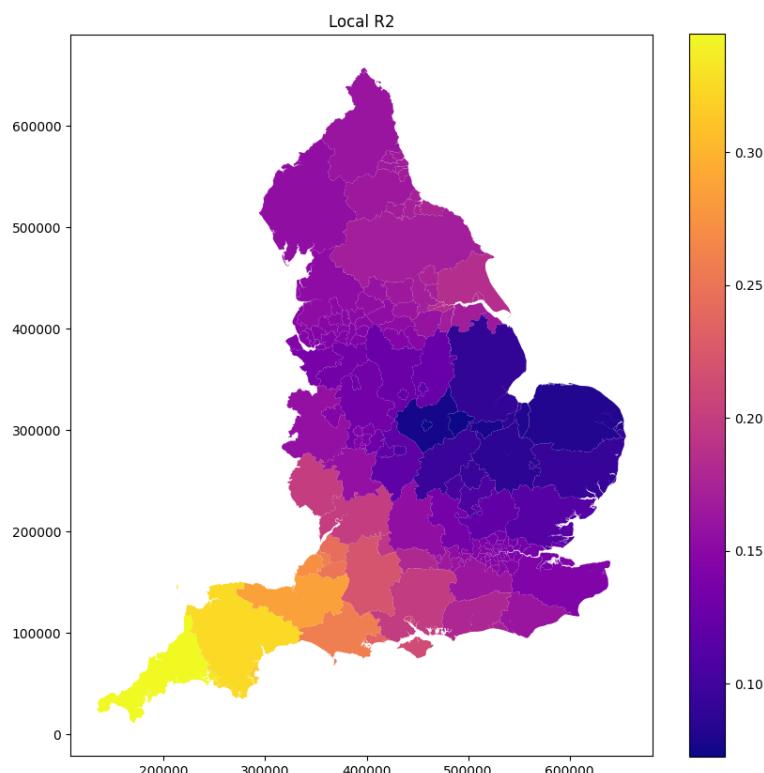
Metric	Value
Bandwidth	113.0000
AICc	793.3382
R2	0.2170
Adj. R2	0.1745

Ringkasan performa model GWR disajikan pada Tabel 4. Dapat dilihat nilai AICc yang lebih rendah dan R^2 yang lebih tinggi dibandingkan model OLS, menunjukkan

peningkatan performa model secara signifikan. Model GWR mampu menjelaskan variasi data lebih baik dengan memperhitungkan aspek lokasi. Bandwidth optimal yang terpilih menunjukkan radius atau jumlah tetangga yang paling sesuai untuk menangkap pola lokal tersebut.

4.4.1 Kebaikan Model Lokal (R^2)

Variasi lokal nilai R^2 (Gambar 5) menunjukkan bahwa kemampuan model dalam menjelaskan variasi kasus COVID-19 berbeda-beda di setiap wilayah.



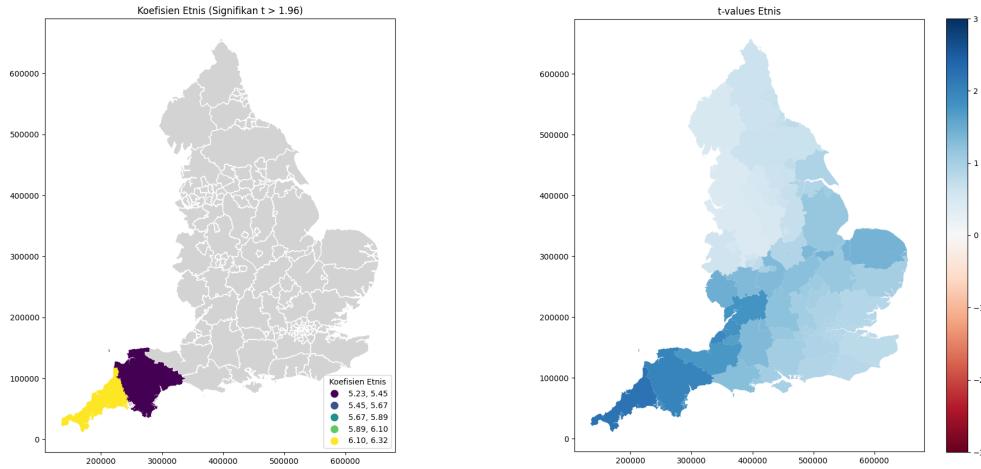
Gambar 5: Peta Local R^2

Peta Local R^2 di atas memperlihatkan bahwa di beberapa wilayah, model GWR memiliki daya jel yang tinggi (warna lebih terang/kuning), sementara di wilayah lain performanya lebih rendah (warna lebih gelap). Hal ini wajar dalam analisis spasial, di mana faktor-faktor penjelas mungkin memiliki kekuatan prediksi yang berbeda tergantung pada konteks lokal masing-masing daerah.

4.4.2 Interpretasi Koefisien Spasial

Berikut adalah peta estimasi parameter lokal (β) dan nilai t-statistik untuk setiap variabel prediktor.

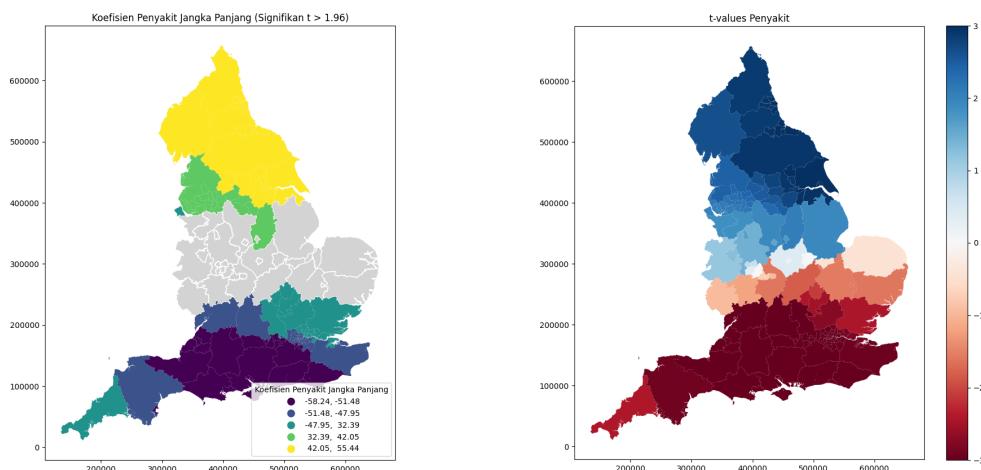
Variabel Etnis Gambar 6 menunjukkan variasi pengaruh proporsi etnis terhadap kasus COVID-19.



Gambar 6: Koefisien Signifikan (kiri) dan t-value (kanan) Variabel Etnis

Peta koefisien etnis (kiri) menunjukkan variasi pengaruh variabel ini. Area yang berwarna abu-abu menunjukkan wilayah di mana koefisien **tidak signifikan** secara statistik ($|t| < 1.96$). Di wilayah yang berwarna (signifikan), terlihat koefisien positif (kuning/hijau) yang mengindikasikan hubungan searah yang kuat. Peta t-value (kanan) memberikan detail kekuatan signifikansi tersebut.

Variabel Penyakit Jangka Panjang Gambar 7 menunjukkan variasi pengaruh penyakit jangka panjang.

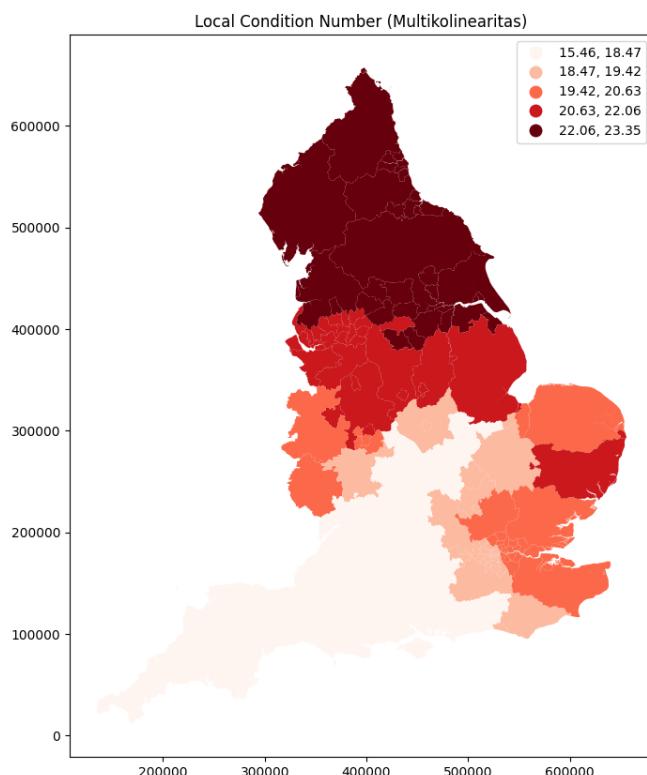


Gambar 7: Koefisien Signifikan (kiri) dan t-value (kanan) Variabel Penyakit

Serupa dengan etnis, peta koefisien penyakit jangka panjang (Gambar 7) telah di-*masking* untuk hanya menampilkan wilayah yang signifikan.

4.4.3 Diagnostik Multikolinearitas Lokal

Stabilitas estimasi parameter GWR perlu dipastikan dengan memeriksa multikolinearitas lokal.



Gambar 8: Peta Local Condition Number

Gambar 8 menunjukkan sebaran nilai *Condition Number* lokal. Nilai-nilai ini beraada dalam rentang yang aman (di bawah 30), yang mengindikasikan bahwa **tidak terdapat masalah multikolinearitas lokal** yang serius antar variabel prediktor (Etnis dan Penyakit Jangka Panjang) di seluruh wilayah studi. Hal ini menjamin bahwa estimasi koefisien GWR bersifat stabil dan dapat diandalkan.

5 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa:

1. Meskipun uji Breusch-Pagan pada model global tidak menunjukkan heteroskedastisitas yang signifikan ($p\text{-value} > 0.05$), penerapan model GWR terbukti membe-

rikan peningkatan performa yang cukup berarti. Hal ini terlihat dari penurunan nilai AICc (814.5 menjadi 793.3) dan peningkatan R^2 dari 0.019 (OLS) menjadi 0.221 (GWR), yang mengindikasikan bahwa pendekatan lokal mampu menjelaskan variasi data dengan lebih baik dibandingkan pendekatan global.

2. Pengaruh faktor etnisitas dan kesehatan jangka panjang terhadap tingkat kasus COVID-19 bervariasi secara geografis. Visualisasi koefisien lokal GWR memperlihatkan bahwa faktor-faktor risiko ini memiliki dampak yang berbeda-beda tergantung pada lokasi. Hal ini menggarisbawahi pentingnya pendekatan kebijakan kesehatan yang bersifat lokal dan spesifik wilayah, karena intervensi yang seragam mungkin tidak efektif di seluruh area.

Pustaka

- [1] Rey, S. J., Arribas-Bel, D., & Wolf, L. J. (2020). *Geographic Data Science with Python*. Retrieved from https://geographicdata.science/book/notebooks/09_gwr.html.
- [2] Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- [3] Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, 8(6), 269.
- [4] Jordahl, K., et al. (2020). geopandas/geopandas: v0.8.1. Zenodo. <https://doi.org/10.5281/zenodo.3946761>
- [5] Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*.