

Presentasi Tugas 3 TK 1 Analisis Data Spasial

# **ANALISIS PERSENTASE PENDUDUK MISKIN MENURUT KABUPATEN/KOTA DI PROVINSI JAWA TIMUR TAHUN 2022 DENGAN MENGGUNAKAN GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)**

Oleh Kelompok F:

1. Ammar Hanafi - 2206051582
2. Bryan Jonathan - 2206052780
3. Renata Shaula Alfino Ritonga - 2206815812

Universitas Indonesia | Statistika | 2025

**Link Presentasi**

[https://youtu.be/z\\_hlyYoCLpw](https://youtu.be/z_hlyYoCLpw)



# DAFTAR ISI

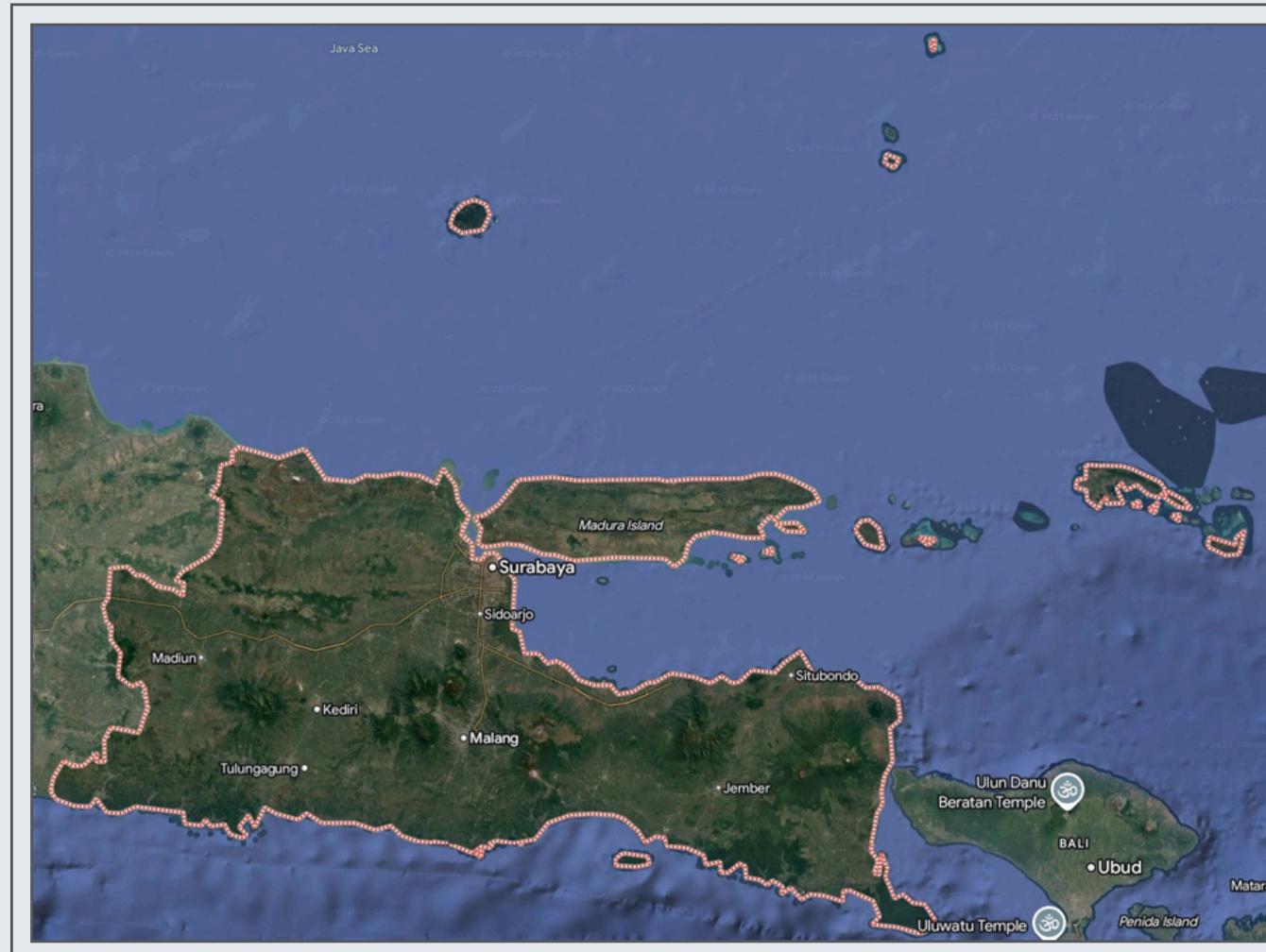
- 01** Cover
- 02** Daftar Isi
- 03** Introduction
- 04** Methods

- 05** Results
- 06** Discussions
- 07** Conclusions
- 08** References



# BAB 1

## Introduction



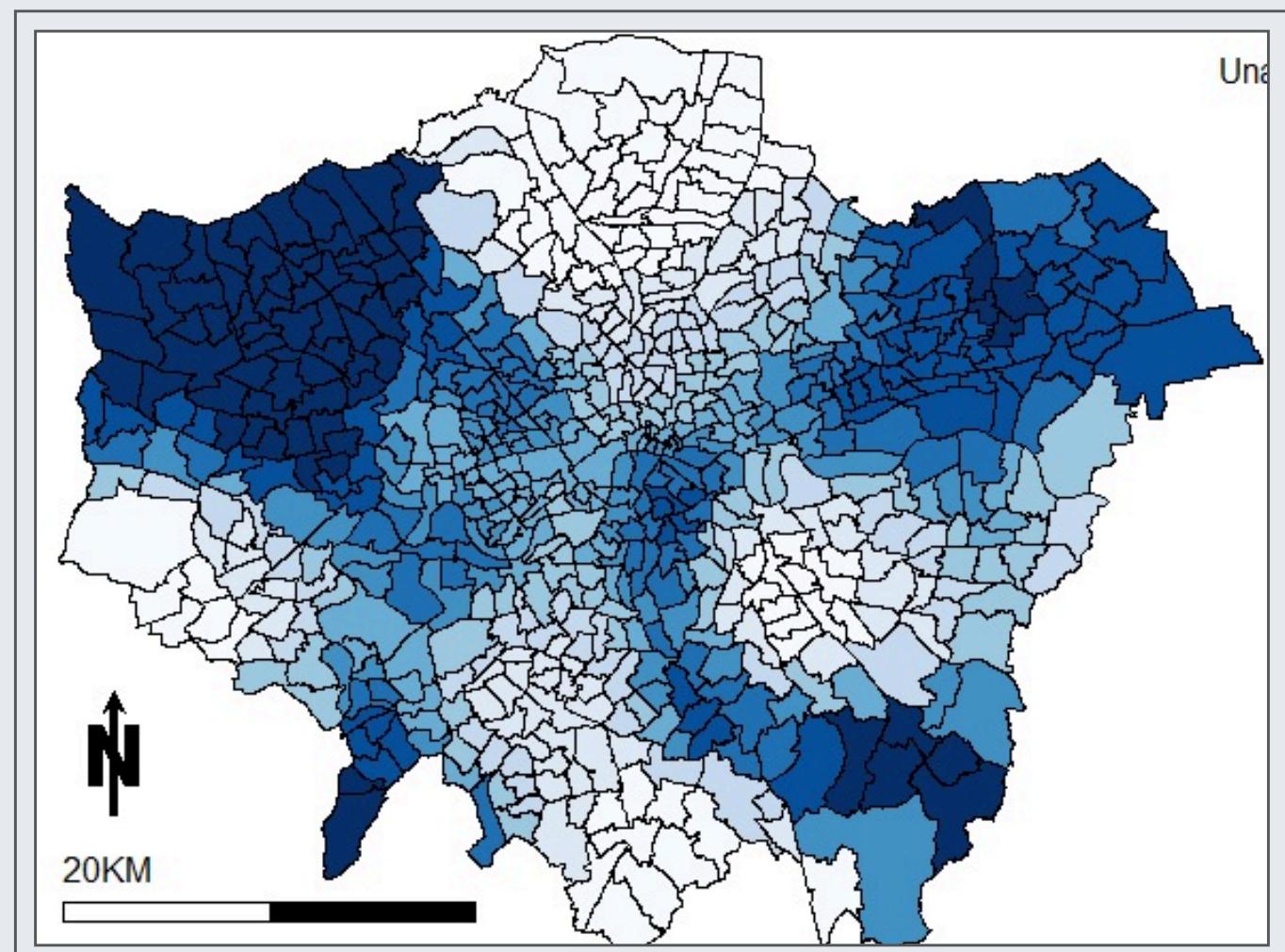
### Pendahuluan Penelitian

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi tingkat kemiskinan di provinsi Jawa Timur.

Hal ini dilakukan dengan memeriksa empat variabel utama: tingkat pengangguran terbuka (TPT), akses ke air minum yang bersih dan layak, harapan lama sekolah, dan jumlah rata-rata rokok yang dikonsumsi oleh individu berusia 30-39 tahun. Menggunakan metode analisis data spasial, penelitian ini berupaya mengungkap hubungan antara variabel-variabel ini dan tingkat kemiskinan di berbagai kabupaten dan kota di Jawa Timur. Dengan memahami pola-pola spasial ini, penelitian ini memberikan wawasan yang dapat membantu menginformasikan strategi pengurangan kemiskinan yang lebih bertarget dan efektif.

# BAB 2

## Methods



# Jenis & Sumber Data

Penelitian ini menggunakan data dari Badan Pusat Statistik (BPS) yang mencakup 38 kabupaten/kota di Provinsi Jawa Timur untuk tahun 2022.

Kabupaten	latitude	longitude	Y	X1	X2	X3	X4
Kabupaten Pacitan	-8,204614	111,08769	13,8	3,65	414	83,95	66,48
Kabupaten Ponorogo	-7,867827	111,466003	9,32	5,51	680	91,50	80,95
Kabupaten Trenggalek	-8,05	111,7166667	10,96	5,37	592	81,23	64,89
...	...	...	...	...	...	...	...
Kota Madiun	-7,629714	111,513702	4,76	6,39	5.514	98,54	68,28
Kota Surabaya	-7,289166	112,734398	4,72	7,62	8.595	95,66	70,53
Kota Batu	-7,8671	112,5239	3,79	8,43	1.116	99,12	70,97

Variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

- Y = Persentase Penduduk Miskin
- X1 = Tingkat Pengangguran Terbuka (TPT)
- X2 = Persentase Rumah Tangga dengan Akses Air Minum
- X3 = Harapan Lama Sekolah
- X4 = Rata-Rata Jumlah Rokok dihabiskan per Minggu oleh Penduduk berumur 30-39



# Multiple Linear Regression

*Multiple Linear Regression* memodelkan hubungan linier antara variabel dependen dan dua atau lebih variabel independen. Bentuk umum dari model ini adalah:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

di mana  $y_i$  adalah variabel dependen,  $x_{ij}$  adalah variabel independen,  $\beta_j$  adalah koefisien regresi, dan  $\varepsilon_i$  adalah error term.

Parameter-parameter tersebut diestimasi dengan menggunakan metode Ordinary Least Squares (OLS), yang meminimumkan jumlah kuadrat kesalahan (Sum of Squared Error/SSE). Signifikansi parameter diuji dengan menggunakan uji-t parsial dan uji F untuk signifikansi secara bersama-sama, sedangkan uji diagnostik digunakan untuk mengevaluasi asumsi-asumsi yang meliputi multikolinieritas, dengan menggunakan VIF; normalitas, dengan menggunakan uji Kolmogorov-Smirnov; autokorelasi dengan menggunakan uji Durbin-Watson; dan homoskedastisitas dengan menggunakan scatterplot.



# Deteksi Heterogenitas Spasial

Untuk menentukan apakah terdapat heterogenitas spasial, digunakan uji Breusch-Pagan dengan hipotesis sebagai berikut:

- H0: Tidak ada heterogenitas spasial
- H1: Terdapat heterogenitas spasial

Hasil yang signifikan dengan nilai p-value  $< 0,05$  mendukung adanya heterogenitas spasial, sehingga memperkuat ketepatan pendekatan *Geographically Weighted Regression*.



# Geographically Weighted Regression (GWR)

GWR adalah bentuk lokal dari regresi linier yang digunakan untuk memodelkan hubungan yang bervariasi secara spasial. Tidak seperti model regresi tradisional yang mengasumsikan hubungan antara variabel dependen dan independen adalah konstan di seluruh ruang, GWR memungkinkan koefisien regresi bervariasi berdasarkan lokasi, yang cocok untuk data dengan heterogenitas spasial. Model GWR untuk observasi ke-i yang terletak pada koordinat  $u_i, v_i$  didefinisikan sebagai:

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

Dimana:

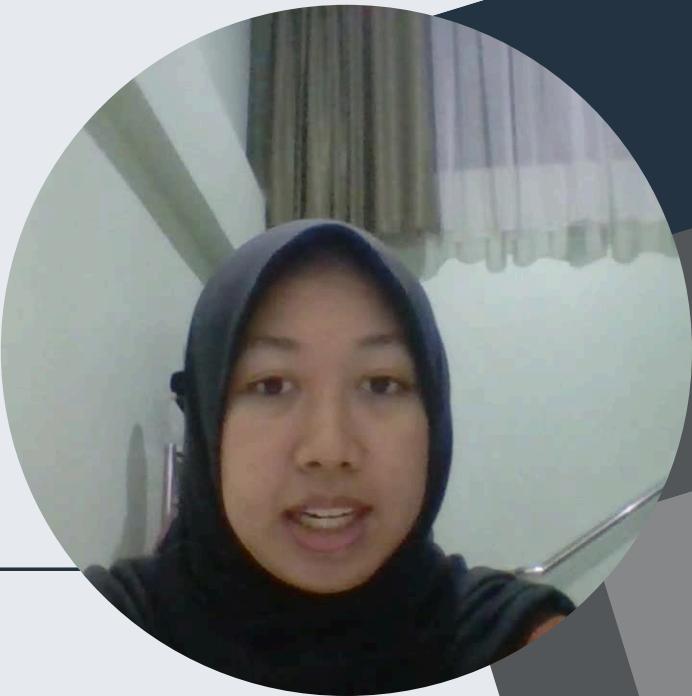
$Y_i$ : nilai variabel dependen pada lokasi ke-i

$X_{ik}$ : nilai variabel independen ke-k pada lokasi ke-i

$(u_i, v_i)$ : koordinat spasial lokasi ke-i

$\beta_k(u_i, v_i)$ : koefisien regresi lokal

$\varepsilon_i$ : error term acak pada lokasi ke-i



# Estimasi Parameter GWR

Vektor koefisien estimasi pada lokasi ke-i,  $u_i, v_i$ , diperoleh dengan meminimumkan Jumlah Kuadrat Galat Tertimbang (JKGT):

$$JKGT = \sum_{i=1}^n w_i(u_i, v_i) \left[ y_i - \beta_0(u_i, v_i) - \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} \right]^2$$

Dalam bentuk matriks, solusinya adalah:

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

Di mana:

X: matriks variabel independen (termasuk intersep)

Y: vektor nilai variabel dependen yang diamati

W( $u_i, v_i$ ): matriks bobot spasial untuk lokasi i, biasanya berupa matriks diagonal



# Fungsi Pembobot Spasial

Bobot dihitung menggunakan fungsi kernel, yang mendefinisikan bagaimana kedekatan mempengaruhi pengaruh titik data yang berdekatan. Kernel yang umum meliputi:

## Gaussian Kernel

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2h^2}\right)$$

## Bisquare Kernel

jika  $d_{ij} < h$ , jika lainnya 0

$$w_{ij} = \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2$$

## Tricube Kernel

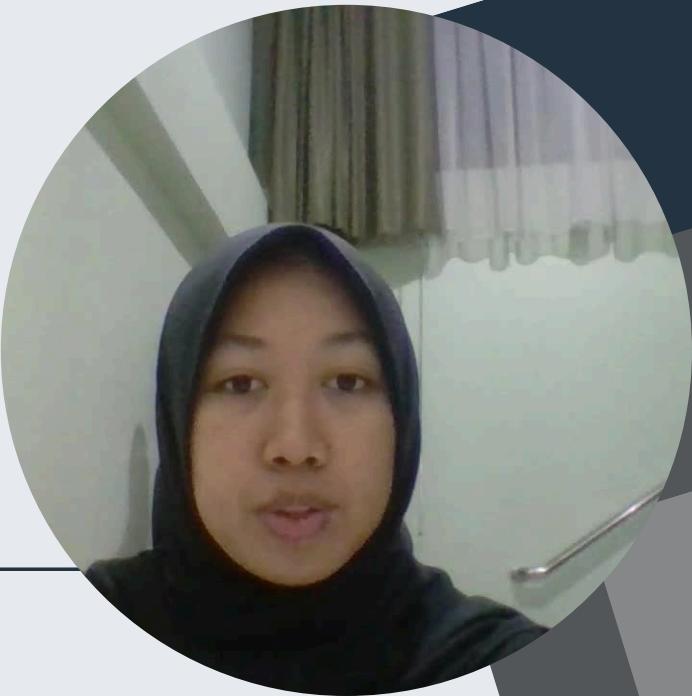
jika  $d_{ij} < h$ , jika lainnya 0

$$w_{ij} = \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3$$

Di mana:

$d_{ij}$ : Jarak Euclidean antara lokasi i dan j

$h$ : parameter *bandwidth* yang mengontrol rentang pengaruh



# Pemilihan *Bandwidth*

*Bandwidth* h sangat penting untuk kinerja model dan biasanya dipilih dengan meminimalkan *Cross-Validation* (CV) atau *Akaike Information Criterion* (AIC). Rumus untuk skor *Cross-Validation* dinyatakan sebagai:

$$CV = \sum_{i=1}^n \left( Y_i - \hat{Y}_{\neq i} \right)^2$$

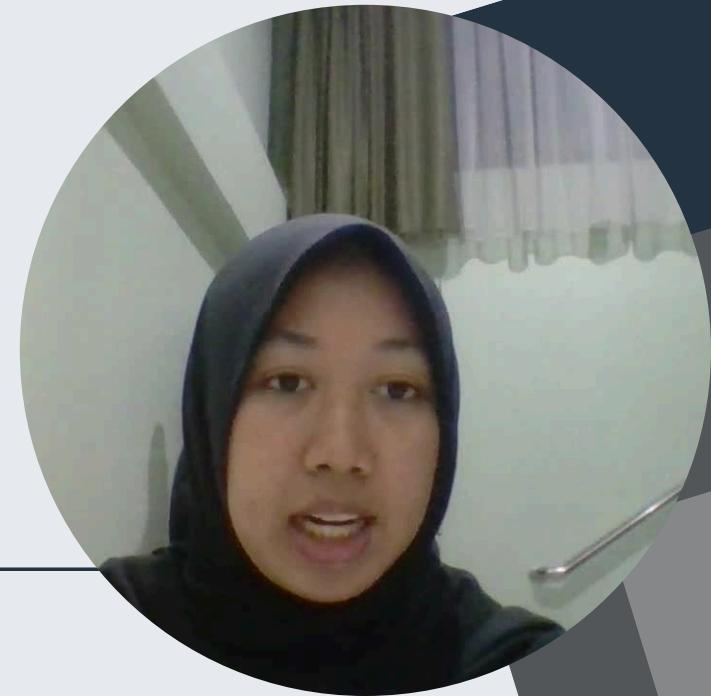
Di mana  $\hat{Y}_i$  adalah nilai fit untuk  $Y_i$  yang diperoleh dengan mengecualikan observasi ke- $i$  selama fitting model.

## Uji Signifikansi Parameter Lokal

Untuk menguji apakah koefisien lokal  $B_k(u_i, v_i)$  secara signifikan berbeda dari nol di lokasi  $i$ , t-statistik dihitung sebagai:

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))}$$

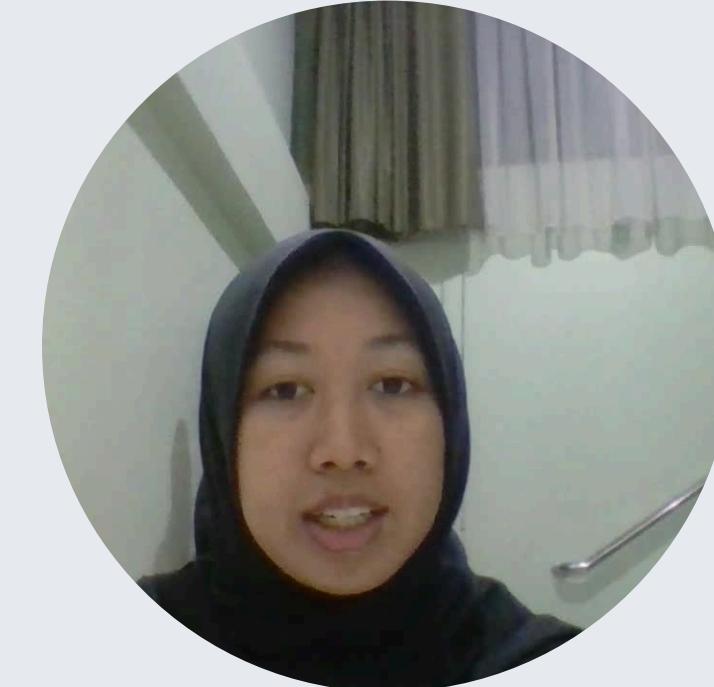
$$SE(\hat{\beta}_k(u_i, v_i)) = \sqrt{C_{kk} \cdot \hat{\sigma}^2}$$



# Evaluasi Model

## a. Akaike Information Criterion (AIC)

$$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n}$$



Di mana:

k: jumlah parameter

n: jumlah observasi

i: error pada observasi ke-i

Nilai AIC yang lebih rendah menunjukkan model yang lebih baik, menyesuaikan dengan kompleksitas model.

## b. Coefficient of Determination (R<sup>2</sup>)

$$R^2 = 1 - \frac{JKG}{JKT}$$

$$JKG = \sum_{i=1}^n \hat{\varepsilon}_i^2; \quad JKT = \sum_{i=1}^n (y_i - \bar{y})^2$$

Di mana:

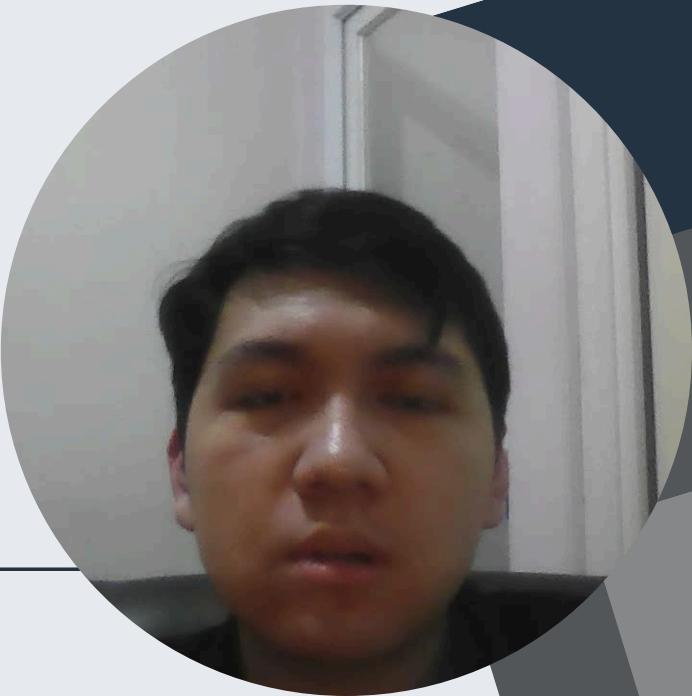
y<sub>i</sub>: data pengamatan ke-i

ybar: rata-rata data pengamatan

Nilai R<sup>2</sup> yang lebih tinggi menunjukkan kekuatan penjelasan yang lebih baik.

# BAB 3

# Hasil





# Analisis Regresi Global

- Dari nilai R-Squared sebesar 0.8083, dapat disimpulkan bahwa model ini mampu menjelaskan sekitar 80.83% variasi pada data.
- Adjusted R-Squared sebesar 0.7851, menunjukkan bahwa meskipun mempertimbangkan jumlah variabel dalam model, tingkat penjelasan masih tetap tinggi.
- F-statistic sebesar 34.79 dengan p-value 2.086e-11 juga mengonfirmasi bahwa model secara keseluruhan signifikan.

Coefficients	Estimate	Standard Error	t-value	
Intercept	48.20624	8.01505	6.014	
X1	-0.86716	0.19862	-4.366	0.000118
X2	-0.20367	0.06923	-2.942	0.005925
X3	-1.86688	0.40005	-4.667	4.91e-05
X4	0.14256	0.02385	5.976	1.03e-06

Residual standard error: 1.98 on 33 degrees of freedom  
 Multiple R-squared: 0.8083, Adjusted R-squared: 0.7851  
 F-statistic: 34.79 on 4 and 33 DF, p-value: 2.086e-11

$$\hat{y} = 48.20624 - 0.86716x_1 - 0.20367x_2 - 1.86688x_3 + 0.14256x_4$$

# Uji Asumsi

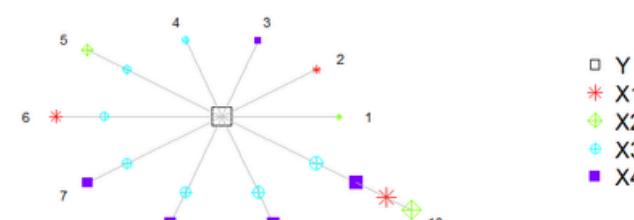
Assumption	Statistical Test	Result	Interpretation
Residual Normality	Shapiro-Wilk Test	$W = 0.99191$ p-value = 0.9934	Residuals are normally distributed ( $p > 0.05$ )
Multicollinearity	Variance Inflation Factor (VIF)	$VIF X_1 = 1.174$ $VIF X_2 = 1.053$ $VIF X_3 = 1.221$ $VIF X_4 = 1.041$	No multicollinearity detected ( $VIF < 5$ )
Residual Autocorrelation	Durbin-Watson Test	$DW = 2.384$ p-value = 0.8331	No autocorrelation detected ( $p > 0.05$ )
Homoscedasticity	Breusch-Pagan Test	$BP = 12.169$ $df = 4$ p-value = 0.01614	Indication of heteroscedasticity ( $p < 0.05$ )

# Pemilihan Peubah Bebas

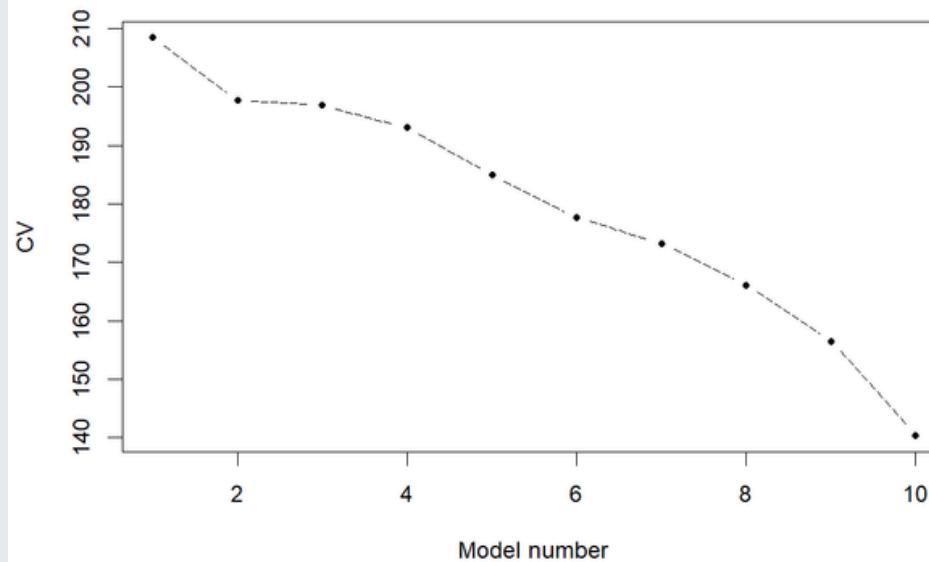
Halaman 16

## Kernel Gaussian

View of GWR model selection with different variables

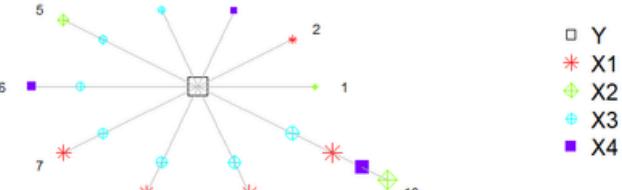


Alternative view of GWR model selection (Gaussian)

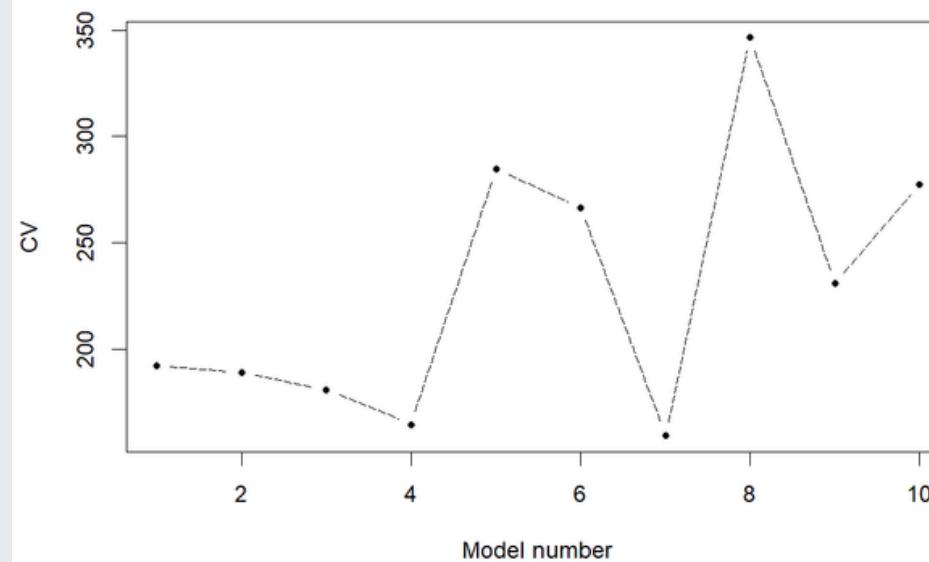


## Kernel Bisquare

View of GWR model selection with different variables

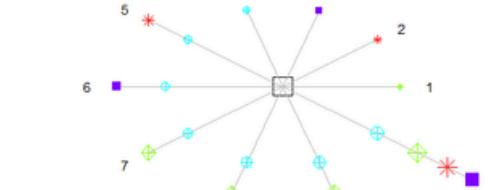


Alternative view of GWR model selection (Bisquare)

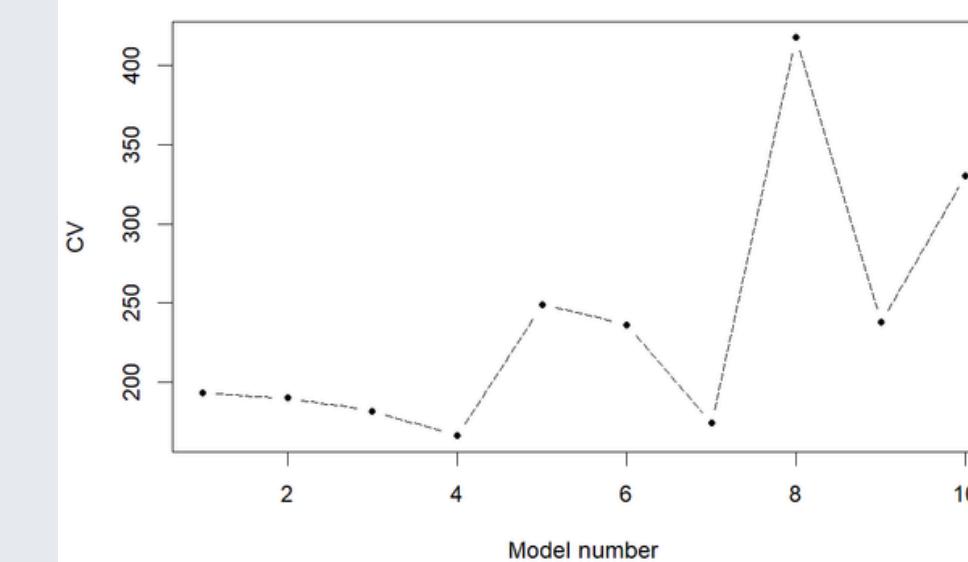


## Kernel Tricube

View of GWR model selection with different variables



Alternative view of GWR model selection (Tricube)



- Untuk Kernel Gaussian, terlihat bahwa nilai CV terus menurun seiring bertambahnya jumlah variabel, menunjukkan bahwa semua variabel berkontribusi dalam model.
- Untuk Kernel Bisquare dan Tricube, pola CV berfluktuasi, tetapi model terbaik tetap tercapai saat seluruh variabel digunakan.
- Oleh karena itu, semua variabel X1, X2, X3, dan X4 akan tetap dimasukkan ke dalam pemodelan GWR selanjutnya.

# Penentuan Bandwidth Optimum

## Kernel Gaussian

Bandwidth optimum untuk fungsi Kernel Gaussian adalah  
6.643368

Bandwidth optimum untuk fungsi Kernel Bisquare adalah  
6.644454

## Kernel Bisquare

## Kernel Tricube

Bandwidth optimum untuk fungsi Kernel Tricube adalah  
6.644414



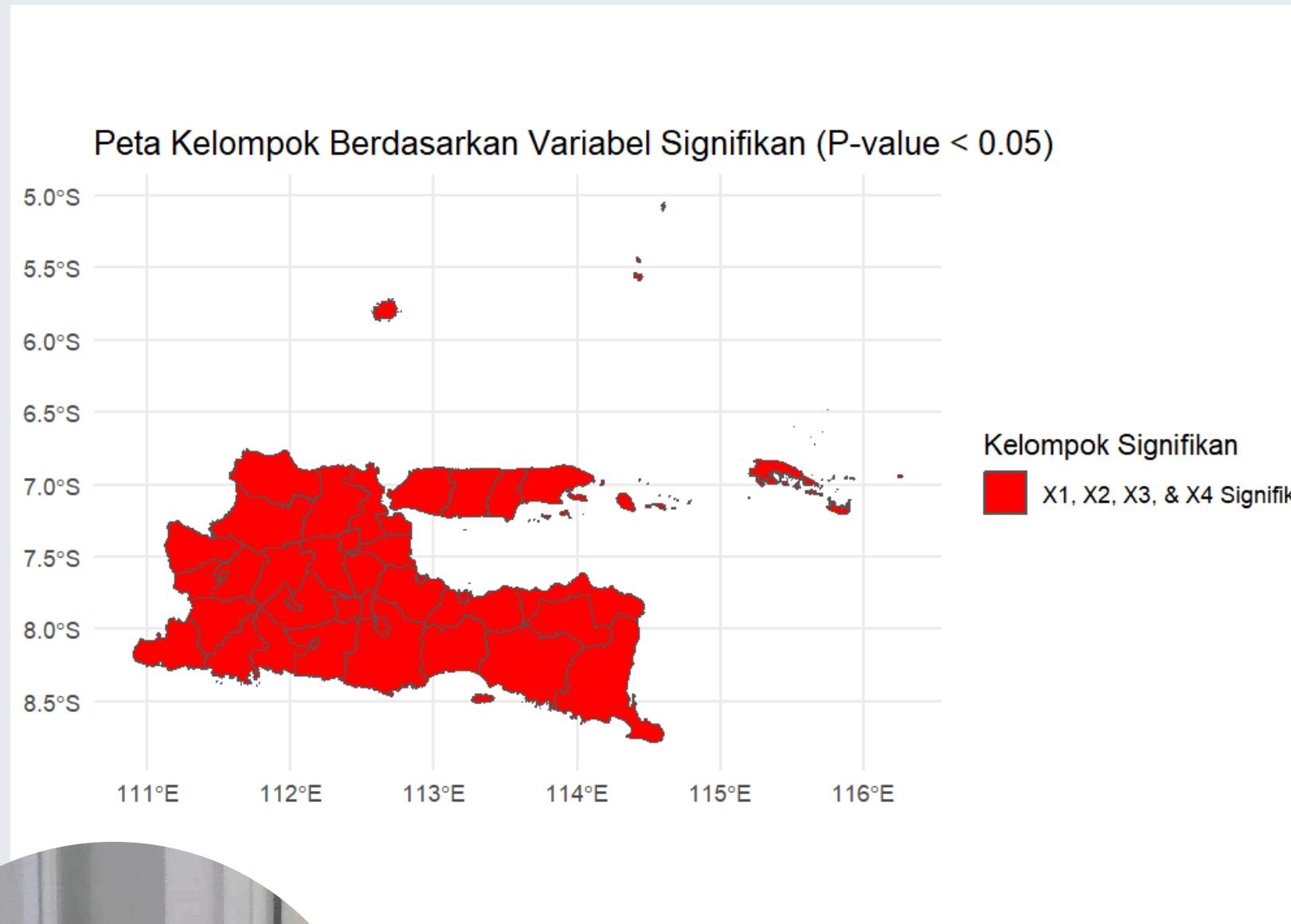
# Pemilihan Model Terbaik

Model	AIC	R-Squared
Global Linear Regression	166.3807	0.8083126
GWR with Gaussian Kernel	158.9220	0.8114638
GWR with Bisquare Kernel	157.2626	0.8225263
GWR with Tricube Kernel	157.9161	0.8181444

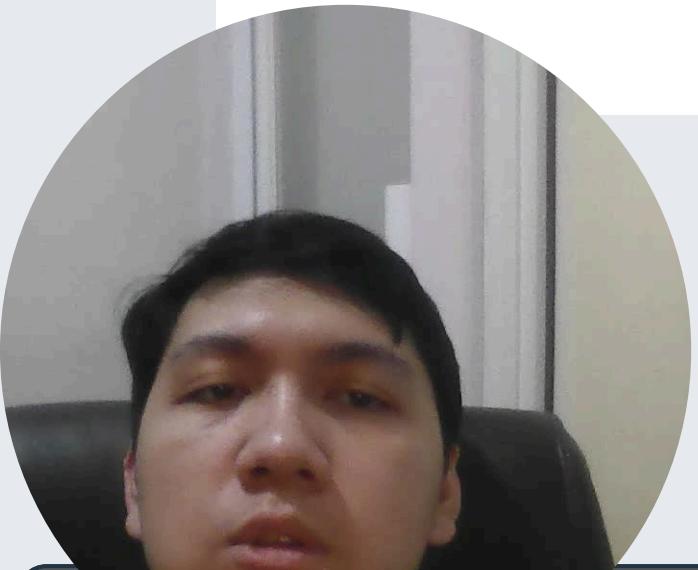
Berdasarkan tabel ini, model GWR dengan Kernel Bisquare memiliki nilai AIC terendah yaitu 157.2626 dan R-Squared tertinggi yaitu 0.8225 dibandingkan model lainnya. Artinya, model GWR Bisquare lebih baik dalam menjelaskan variasi data sekaligus menghasilkan prediksi yang lebih akurat dibandingkan regresi global maupun GWR dengan kernel lain. Sehingga, model GWR dengan Kernel Bisquare dipilih sebagai model terbaik.



# Sebaran Peubah Signifikan

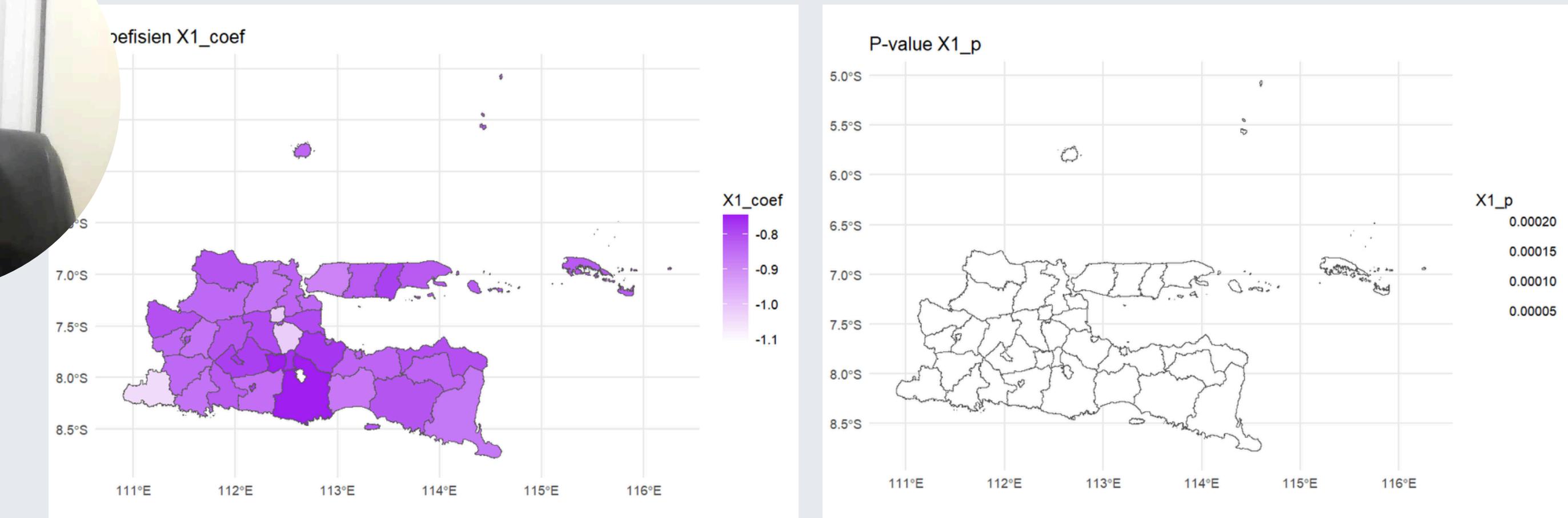
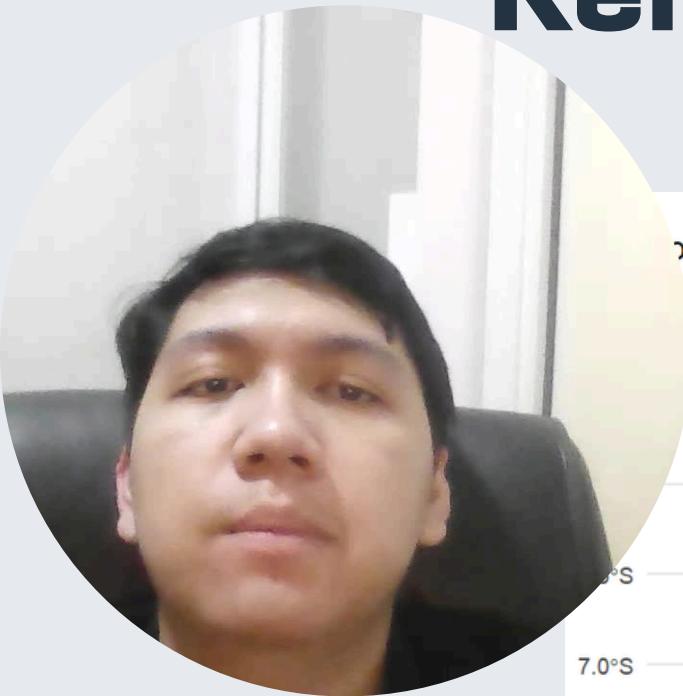


Pada peta ini terlihat bahwa semua wilayah di Jawa Timur berwarna merah, yang berarti keempat variabel  $X_1$ ,  $X_2$ ,  $X_3$ , dan  $X_4$  berpengaruh signifikan terhadap persentase penduduk miskin. Ini menunjukkan bahwa hubungan antar variabel tidak hanya terjadi di sebagian wilayah, tetapi merata secara spasial di seluruh provinsi Jawa Timur.



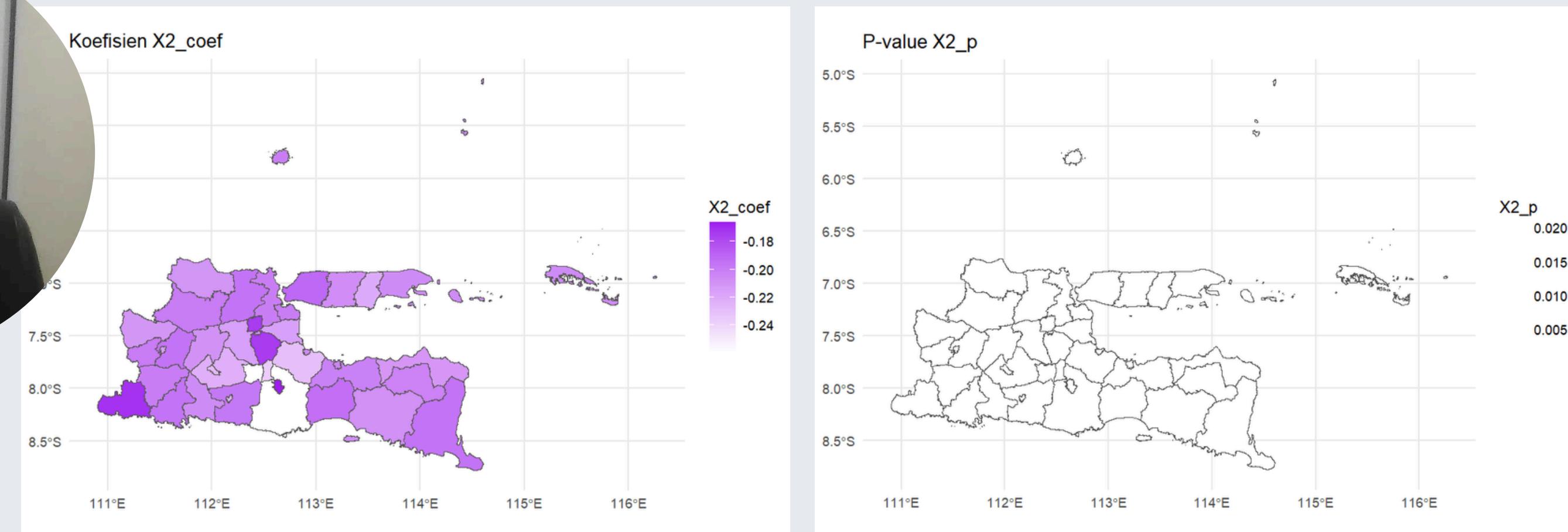
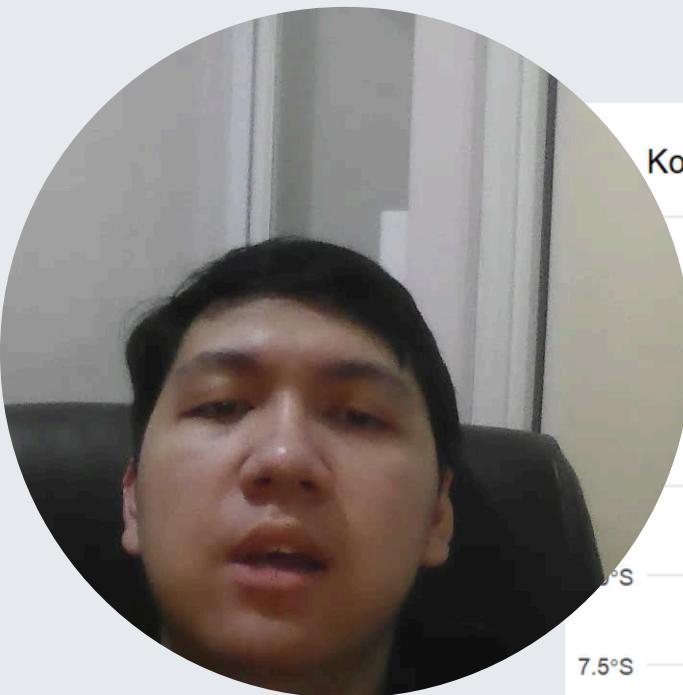
# Keragaman Spasial Penduga Parameter Tingkat Pengangguran Terbuka (TPT)

Halaman 20



- Peta ini menunjukkan bahwa tingkat pengangguran terbuka (TPT) memiliki hubungan negatif terhadap persentase penduduk miskin.
- Semakin tinggi tingkat pengangguran terbuka, justru persentase kemiskinan cenderung menurun.
- Wilayah dengan warna ungu lebih gelap, seperti Kabupaten Malang, Jember, dan Kediri, menunjukkan pengaruh negatif yang lebih kuat.
- Berdasarkan peta p-value, seluruh wilayah signifikan secara statistik.

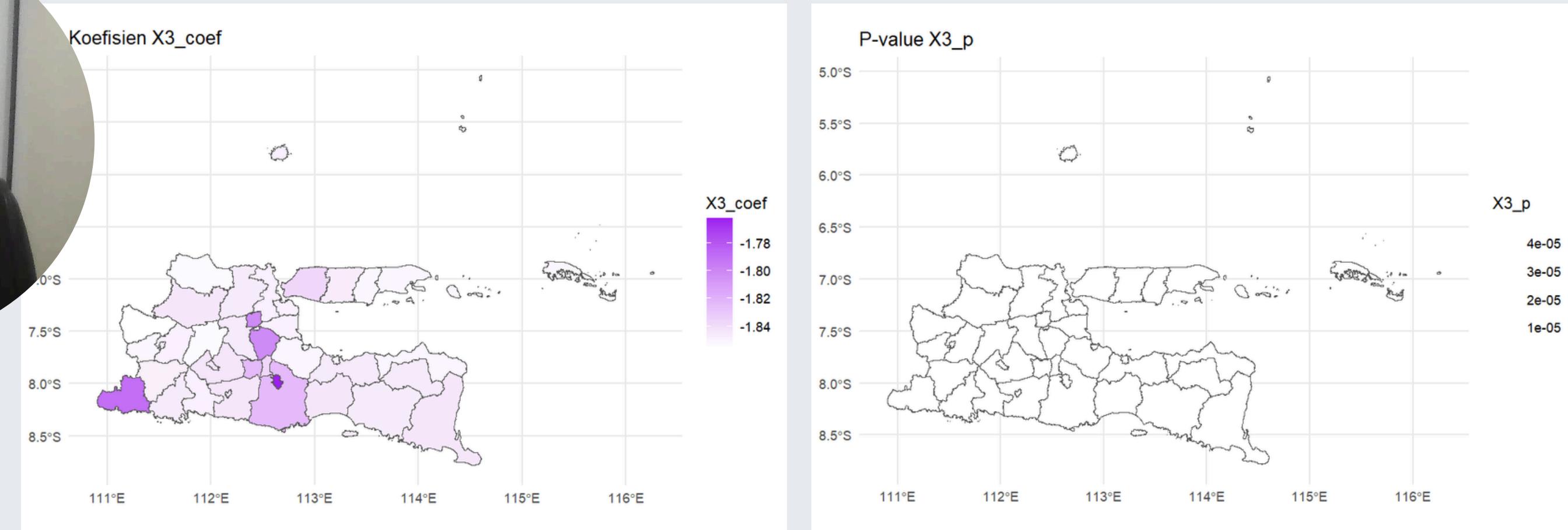
# Keragaman Spasial Penduga Parameter Persentase Rumah Tangga dengan Akses Air Minum



- Peta ini menunjukkan bahwa akses air minum layak memiliki hubungan negatif terhadap persentase penduduk miskin.
- Artinya, semakin banyak rumah tangga yang memiliki akses air minum layak, semakin rendah tingkat kemiskinan.
- Wilayah dengan warna ungu lebih gelap, seperti Pacitan, Lumajang, dan Ponorogo, menunjukkan pengaruh negatif yang lebih kuat.
- Seluruh wilayah juga signifikan berdasarkan peta p-value.

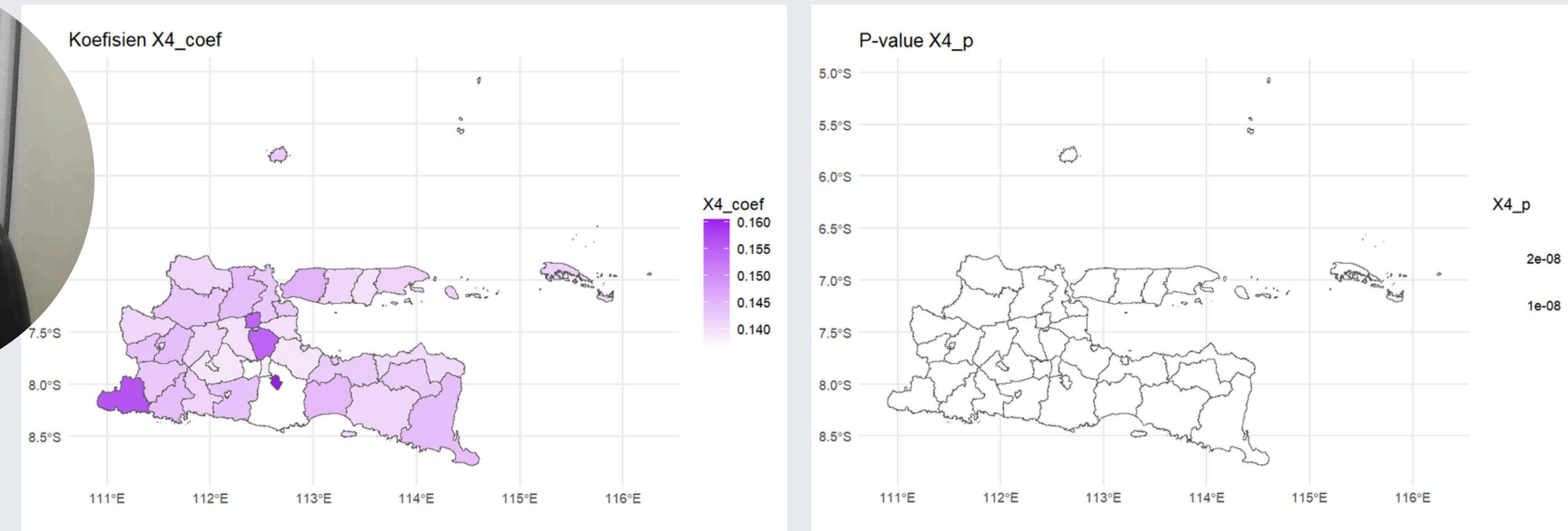
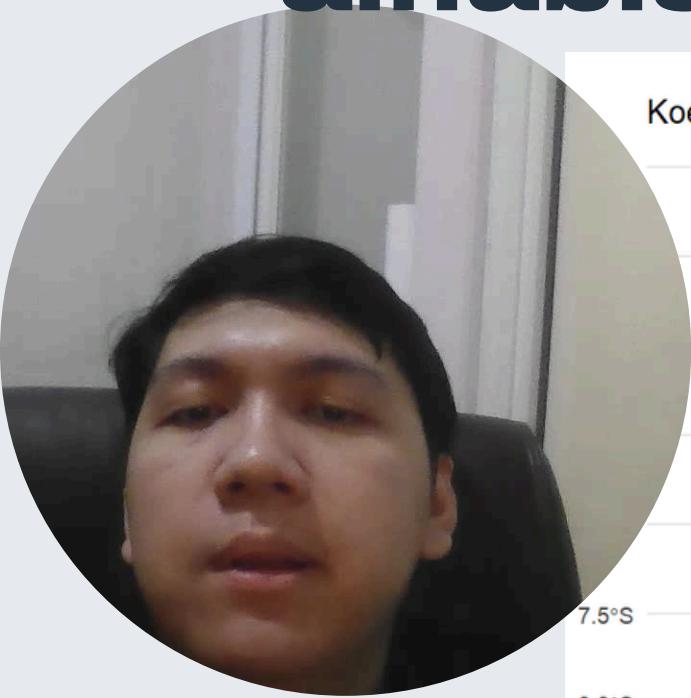
# Keragaman Spasial Penduga Parameter Harapan Lama Sekolah

Halaman 22



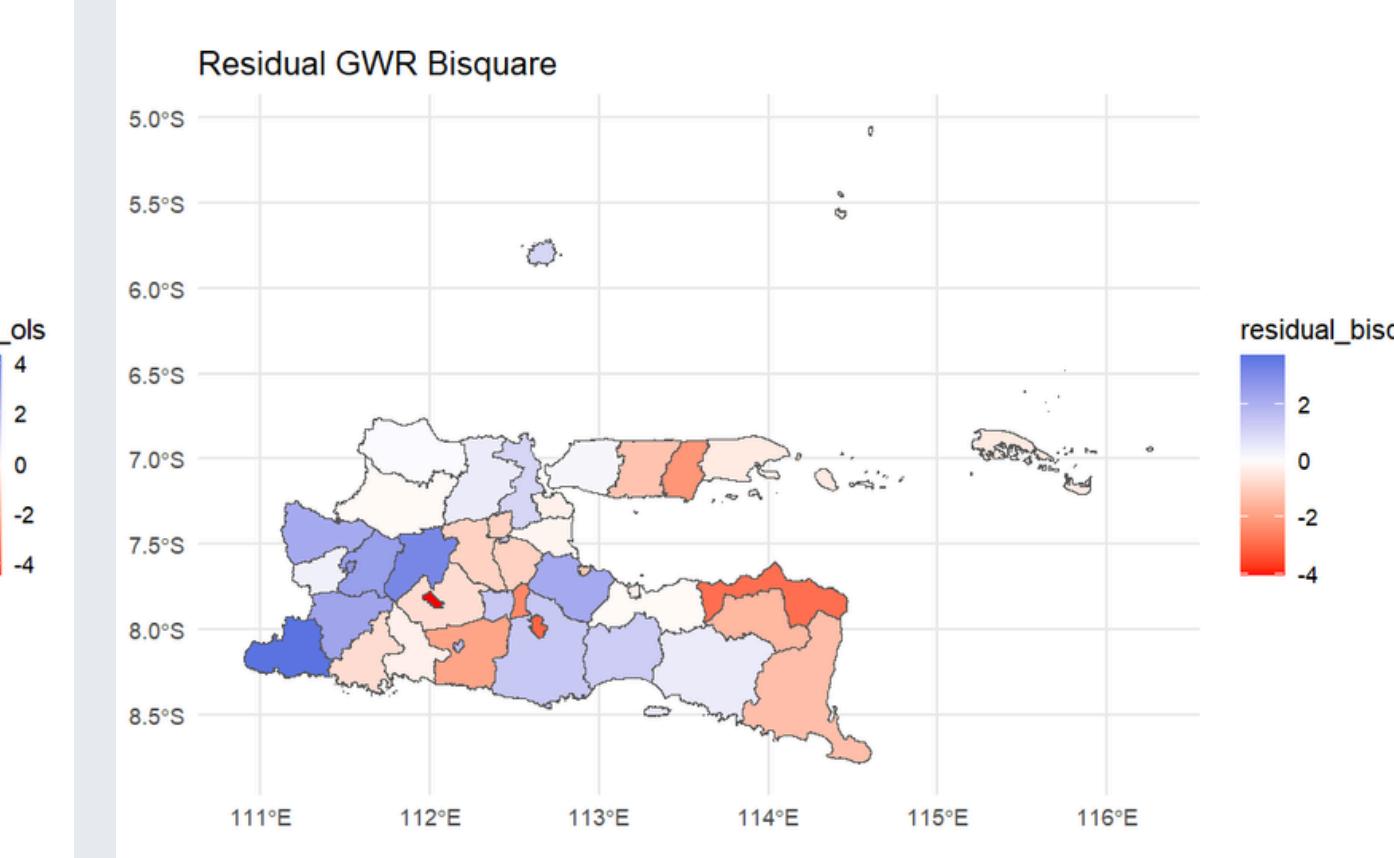
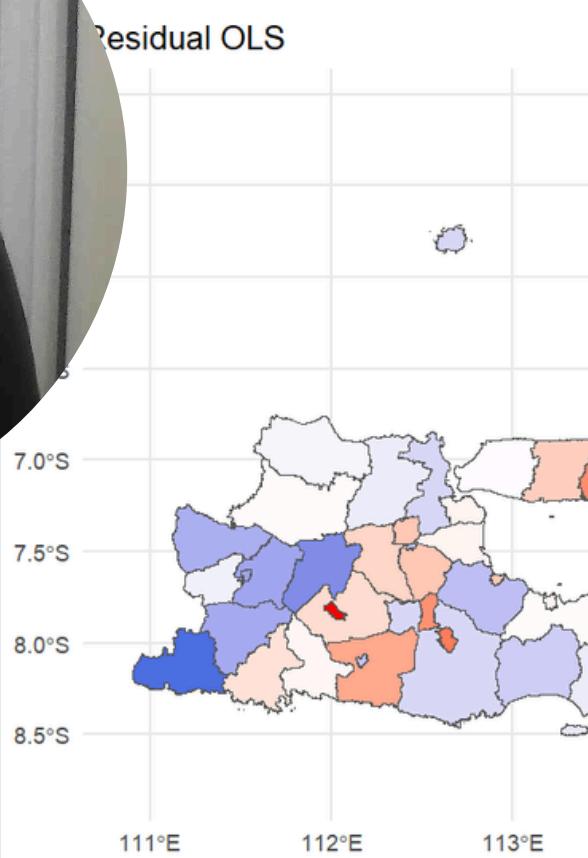
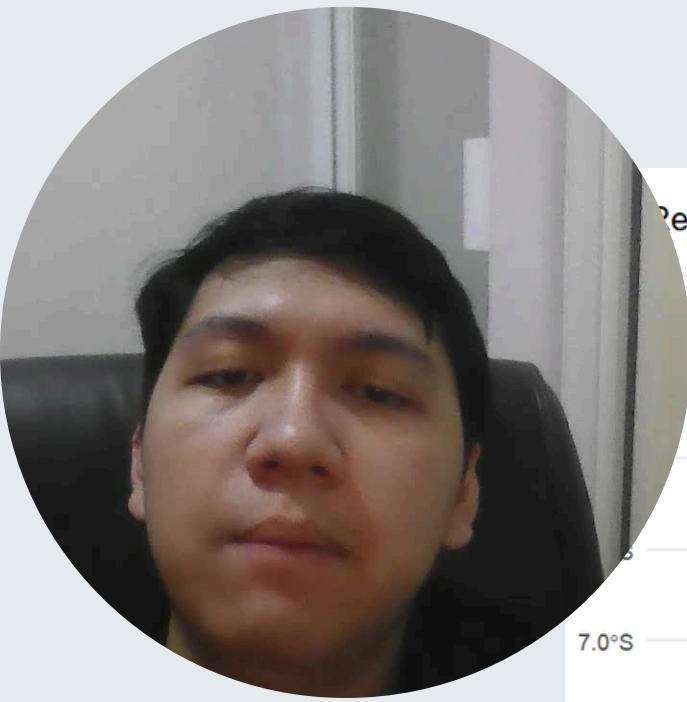
- Peta ini menunjukkan bahwa harapan lama sekolah memiliki hubungan negatif terhadap persentase penduduk miskin.
- Semakin tinggi harapan lama sekolah, semakin rendah tingkat kemiskinan.
- Wilayah seperti Pacitan, Ponorogo, dan Blitar terlihat memiliki pengaruh negatif yang lebih kuat.
- Berdasarkan peta p-value, seluruh wilayah signifikan secara statistik.

# Keragaman Spasial Penduga Rata-Rata Jumlah Rokok dihabiskan per Minggu oleh Penduduk berumur 30-39



- Peta ini menunjukkan bahwa konsumsi rokok pada penduduk usia 30-39 tahun memiliki hubungan positif terhadap persentase penduduk miskin.
- Semakin tinggi konsumsi rokok, semakin tinggi tingkat kemiskinan.
- Wilayah seperti Pacitan, Ponorogo, dan Blitar memperlihatkan pengaruh positif yang lebih kuat.
- Berdasarkan peta p-value, seluruh wilayah signifikan secara statistik.

# Keragaman Spasial Residual



Model	Min	Max
Model Global	-4.479115	4.326362
Model GWR Gaussian	-4.063486	3.712704

- Pada peta residual, terlihat bahwa model GWR Bisquare menghasilkan sebaran residual yang lebih kecil dibandingkan model global.
- Nilai residual GWR Bisquare berkisar antara -4.06 hingga 3.71, sedangkan residual model global lebih besar, dari -4.48 hingga 4.32.
- Ini menunjukkan bahwa model GWR Bisquare lebih akurat dalam menangkap variasi lokal dibandingkan model global.

# Bab 4

## Discussions

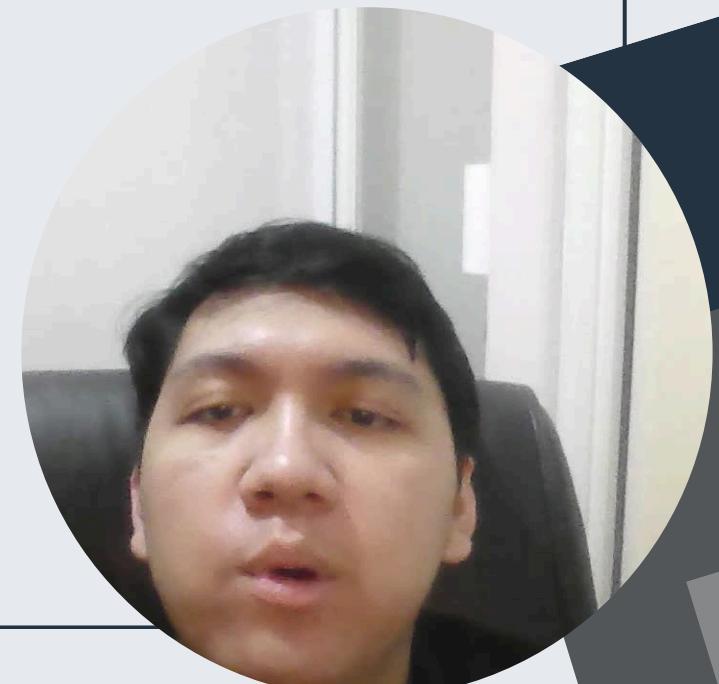
Analisis tingkat kemiskinan di Jawa Timur mengungkapkan korelasi yang signifikan dengan beberapa faktor sosial ekonomi. Tingkat pengangguran yang tinggi (TPT) sangat terkait dengan kemiskinan yang lebih tinggi, terutama di daerah perkotaan. Demikian pula, daerah dengan akses terbatas ke air bersih dan pencapaian pendidikan yang lebih rendah cenderung memiliki tingkat kemiskinan yang lebih tinggi, menekankan pentingnya layanan dasar dan pendidikan dalam pengurangan kemiskinan. Selain itu, tingkat merokok yang lebih tinggi di bidang tertentu menunjukkan hubungan antara perilaku kesehatan dan kesejahteraan ekonomi. Temuan ini menyoroti perlunya intervensi yang ditargetkan untuk mengatasi faktor-faktor ini dan mengurangi kemiskinan di Jawa Timur.

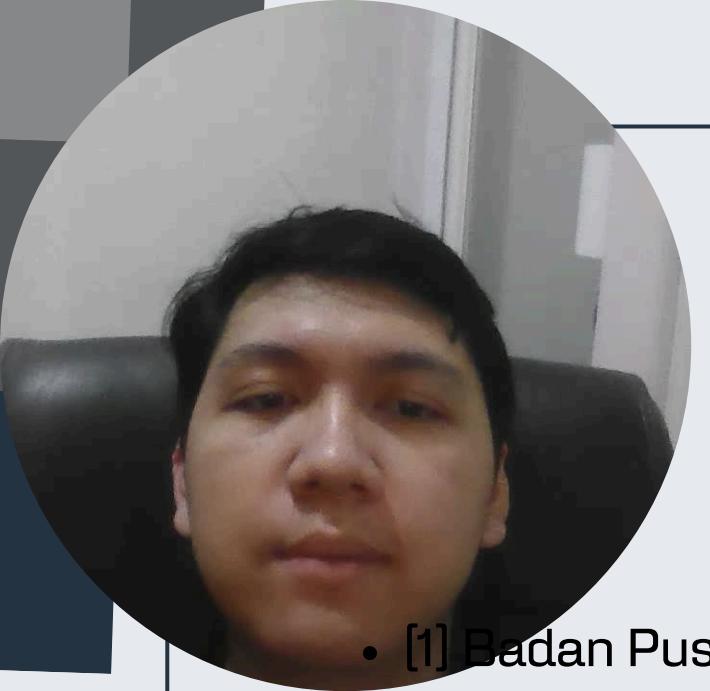


# Bab 5

## Conclusions

- Hasil analisis menunjukkan bahwa seluruh variabel, yaitu tingkat pengangguran terbuka, akses air minum layak, harapan lama sekolah, dan konsumsi rokok, berpengaruh signifikan terhadap persentase penduduk miskin di Jawa Timur.
- Model GWR dengan kernel Bisquare dipilih sebagai model terbaik karena menghasilkan nilai AIC terendah, R-Squared tertinggi, dan residual yang lebih kecil dibandingkan model regresi global.
- Hal ini membuktikan bahwa GWR Bisquare lebih mampu menangkap variasi lokal dalam data kemiskinan, sehingga memberikan hasil analisis yang lebih akurat dan sesuai untuk wilayah Jawa Timur.





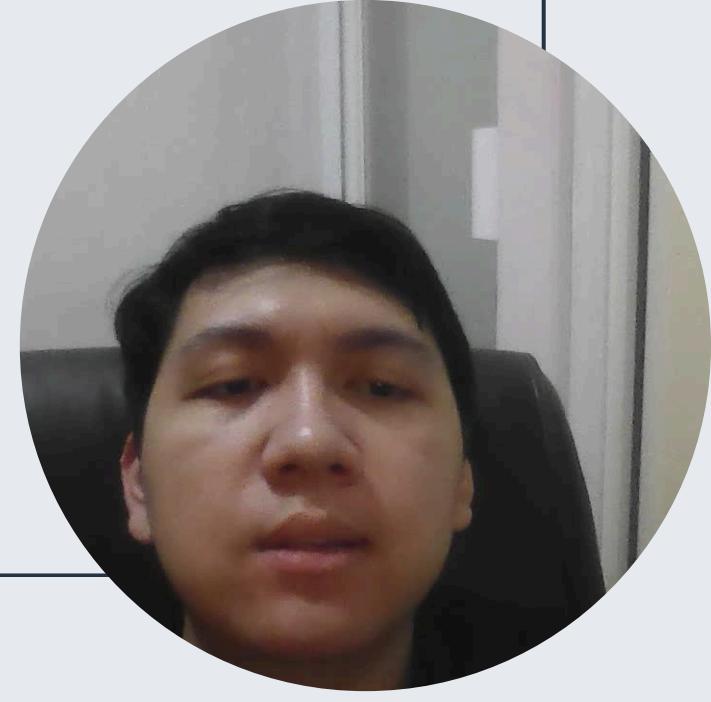
# Bab 6

## Reference

- [1] Badan Pusat Statistik Jawa Timur, "Tingkat Pengangguran Terbuka (TPT) menurut Kabupaten/Kota," BPS Jawa Timur, 2022. [Online]. Available: <https://jatim.bps.go.id/statistics-table/2/NTQjMg==/tingkat-pengangguran-terbuka-tpt-menurut-kabupaten-kota.html> [Accessed: 27 April 2025].
- [2] Badan Pusat Statistik Jawa Timur, "Percentase Rumah Tangga Menurut Kabupaten/Kota Sumber Air Minum Bersih dan Akses Air Minum Layak," BPS Jawa Timur, 2022. [Online]. Available: <https://jatim.bps.go.id/statistics-table/1/MjkwMCMx/persentase-rumah-tangga-menurut-kabupaten-kota-sumber-air-minum-bersih-dan-akses-air-minum-layak-2022.html> [Accessed: 27 April 2025].
- [3] Badan Pusat Statistik Jawa Timur, "Harapan Lama Sekolah," BPS Jawa Timur, 2022. [Online]. Available: <https://jatim.bps.go.id/statistics-table/2/NDY4IzI=/harapan-lama-sekolah.html> [Accessed: 27 April 2025].
- [4] Badan Pusat Statistik Jawa Timur, "Rata-rata Jumlah Batang Rokok per Minggu yang Dihisap Penduduk Usia 5 Tahun ke Atas yang Merokok Tembakau Sebulan Terakhir di Jawa Timur Menurut Kabupaten/Kota Kelompok Umur," BPS Jawa Timur, 2022. [Online]. Available: <https://jatim.bps.go.id/statistics-table/1/Mjc4MiMx/rata-rata-jumlah-batang-rokok-per-minggu-yang-dihisap-penduduk-usia-5-tahun-ke atas-yang-merokok-tebakau-sebulan-terakhir-di-jawa-timur-menurut-kabupaten-kota-kelompok-umur-2022.html> [Accessed: 27 April 2025].
- [5] Lutfiani, N., Sugiman, S., & Mariani, S. "Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-Square." UNNES Journal of Mathematics, 8(1), pp. 82-91, 2019.

# Terima Kasih

Kelompok F | Universitas Indonesia | Statistika | 2025



# Lampiran

[Google Drive Link](#)