



# **ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KEMISKINAN DI PROVINSI JAWA TIMUR TAHUN 2024: PENDEKATAN PERBANDINGAN MODEL**

**Tugas 4 Topik Khusus 1: Spasial  
Kelompok F**

**Link Presentasi:**

**<https://youtu.be/ODEhQMngCi8>**



# **Anggota**

## **Kelompok F**



**Ammar Hanafi – 2206051582**

**Bryan Jonathan – 2206052780**

**Renata Shaula Alfino Ritonga –  
2206815812**



# Daftar Isi

01 Pendahuluan

02 Landasan Teori

03 Data & Metode

04 Hasil & Pembahasan

05 Kesimpulan & Saran







# 01 PENDAHULUAN



# Latar Belakang



Kemiskinan merupakan isu sosial-ekonomi yang kompleks dan masih menjadi tantangan utama di Indonesia, termasuk di Provinsi Jawa Timur. Berdasarkan data Badan Pusat Statistik (BPS), persentase penduduk miskin di Jawa Timur pada Maret 2024 tercatat sebesar 9.79%, dengan jumlah penduduk miskin mencapai 3.983 juta orang. Meskipun terdapat penurunan dibandingkan tahun sebelumnya, disparitas kemiskinan antar wilayah masih signifikan. Sebagai contoh, Kabupaten Sampang mencatatkan angka persentase penduduk miskin sebesar 20.83% pada tahun 2024.

Fenomena kemiskinan di Jawa Timur dapat dilihat dari kesenjangan yang ada antara wilayah-wilayah yang lebih berkembang dengan yang kurang berkembang. Di wilayah yang lebih maju, umumnya tingkat kemiskinan lebih rendah dan terdapat lebih banyak akses terhadap berbagai layanan dasar seperti pendidikan, kesehatan, dan infrastruktur. Sebaliknya, di wilayah yang lebih terisolasi, tingkat kemiskinan cenderung lebih tinggi, yang berhubungan dengan kurangnya akses terhadap layanan-layanan tersebut.



# Tujuan Penelitian

①

Menganalisis pengaruh variabel-variabel independen terhadap tingkat kemiskinan di Provinsi Jawa Timur tahun 2024.

②

Membandingkan kinerja model regresi spasial dan Random Forest dalam memprediksi tingkat kemiskinan berdasarkan faktor-faktor tersebut.







# 02 LANDASAN TEORI





# Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari dua jenis, yaitu variabel dependen (Y) dan variabel independen (Xi). Berikut merupakan definisi dari tiap variabel yang digunakan.

Variabel	Definisi
Persentase penduduk miskin (Y)	Persentase penduduk yang hidup di bawah garis kemiskinan nasional.
Gini Ratio (X1)	Ukuran tingkat kesenjangan pembagian pendapatan relatif antar penduduk suatu wilayah.
Rata-rata Lama Sekolah (X2)	Jumlah tahun belajar penduduk usia 15 tahun ke atas yang telah diselesaikan dalam pendidikan formal.
Jumlah Tenaga Medis (X3)	Jumlah dokter, dokter gigi, dokter spesialis, dan tenaga medis lainnya yang memberikan pelayanan kesehatan di suatu wilayah.
Jumlah Sepeda Motor (X4)	Jumlah sepeda motor yang terdaftar di suatu wilayah administratif.





# Regresi Linier

Regresi linier berganda adalah metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen dengan dua atau lebih variabel independen secara simultan.

Model regresi linier berganda untuk  $n$  observasi dan  $p$  variabel independen dapat digambarkan sebagai berikut:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \dots + \beta_p x_{1p} + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \dots + \beta_p x_{2p} + \epsilon_2 \\ y_3 &= \beta_0 + \beta_1 x_{31} + \dots + \beta_p x_{3p} + \epsilon_3 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \dots + \beta_p x_{np} + \epsilon_n \end{aligned}$$

Secara matriks, model ini dapat dituliskan sebagai:

$$Y = X\beta + \epsilon$$

Estimasi parameter yang diperoleh menggunakan metode Ordinary Least Square (OLS) adalah:

$$\hat{\beta} = (X'X)^{-1}X'Y$$



# Regresi Linier



## Uji Signifikansi Simultan

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$H_1$ : minimal terdapat satu  $\beta_k \neq 0$ ;  $k = 1, 2, \dots, p$ .

Statistik ujinya adalah:

$$F = \frac{\frac{(SS_{yy} - SSR)}{p}}{\frac{SSR}{[n - (p + 1)]}}$$

$H_0$  ditolak pada taraf signifikansi  $\alpha$  jika  $F > F(\alpha, p, n - (p + 1))$

## Uji Signifikansi Parsial

$$H_0: \beta_j = 0, j = 1, 2, \dots, p$$

$$H_1: \beta_j \neq 0, j = 1, 2, \dots, p$$

Statistik ujinya adalah:

$$t = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}$$

$H_0$  ditolak pada taraf signifikansi  $\alpha$  jika  $|t| > t_{\alpha/2}$ .



# Uji Asumsi Regresi Linier



## Uji Normalitas

Uji normalitas dapat dilakukan dengan uji Anderson-Darling.

$H_0$  : Residual berdistribusi normal

$H_1$ : Residual tidak berdistribusi normal

Statistik ujinya adalah:

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} \left[ \ln F(Y_i) + \ln(1 - F(Y_{n+1-i})) \right]$$

$H_0$  ditolak pada taraf signifikansi  $\alpha$  jika  $p\text{-value} < \alpha$ .

## Uji Autokorelasi

Uji autokorelasi dapat dilakukan dengan uji Durbin-Watson.

$H_0 : \rho = 0$  (tidak terdapat autokorelasi)

$H_1 : \rho \neq 0$  (terdapat autokorelasi)

$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}$$

$H_0$  ditolak jika  $p\text{-value} < \alpha$ .

## Uji Multikolinieritas

Multikolinearitas dapat dilihat dengan nilai Variance Inflation Factor (VIF). Nilai  $VIF > 10$  menandakan bahwa terdapat multikolinearitas.

$$VIF = \frac{1}{(1 - R_k^2)}$$





# **03 DATA DAN METODE**





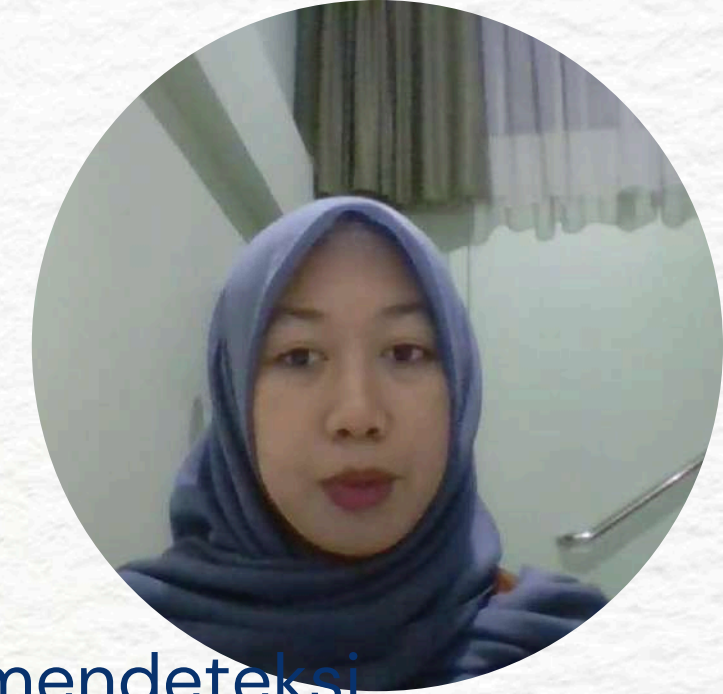
# Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur untuk tahun 2024. Data ini mencakup tingkat kemiskinan (persentase penduduk miskin) dan beberapa variabel independen yang dipilih berdasarkan teori dan relevansi terhadap tingkat kemiskinan berdasarkan kabupaten/kota. Berikut adalah cuplikan 5 data pertama.

1	Kabupaten	latitude	longitude	Y	X1	X2	X3	X4
2	Pacitan	-8,204614	111,08769	13,08	0,327	7,9	47	187184
3	Ponorogo	-7,867827	111,466003	9,11	0,326	7,8	59	472695
4	Trenggalek	-8,05	111,7166667	10,5	0,350	7,92	58	311129
5	Tulungagung	-8,0666667	111,9	6,28	0,321	8,68	53	689732
6	Blitar	-8,1014419	112,162762	8,16	0,353	7,87	35	493701



# Uji Keragaman Spasial



## Uji Moran's I

Indeks Moran I berada pada rentang -1 hingga +1 dan dapat digunakan untuk mendeteksi autokorelasi spasial.

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Hipotesis untuk mendeteksi dependensi spasial:

$H_0 : I = 0$  (Tidak ada autokorelasi antar lokasi)

$H_1 : I \neq 0$  (Terdapat autokorelasi antar lokasi)

Statistik ujinya adalah:

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

$H_0$  ditolak pada taraf signifikansi  $\alpha$  jika  $|Z(I)| > Z_{\alpha/2}$ .

## Uji Keragaman Spasial

Uji yang digunakan untuk mendeteksi keragaman spasial adalah uji Breusch-Pagan.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \sigma^2 \text{ (varians antar lokasi adalah sama)}$$

$$H_1 : \sigma_i^2 \neq \sigma^2 \text{ (terdapat minimal satu varians yang berbeda)}$$

Statistik ujinya adalah:

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f$$

$H_0$  ditolak jika  $BP > \chi^2(p)$ .



# Uji Keragaman Spasial



## Lagrange Multiplier Lag

Digunakan untuk melihat apakah terdapat dependensi spasial pada lag.

$H_0: \rho = 0$  (tidak ada ketergantungan spasial pada variabel dependen)

$H_1: \rho \neq 0$  (terdapat ketergantungan spasial pada variabel dependen)

Statistik ujinya adalah:

$$LM_{\rho} = \frac{\left[ \frac{e'WY}{\left(\frac{ee'}{n}\right)} \right]^2}{D}$$

$H_0$  ditolak jika  $LM_{\rho} > \chi^2_{\alpha(p)}$

## Lagrange Multiplier Error

Digunakan untuk melihat apakah terdapat dependensi spasial pada error.

$H_0: \lambda = 0$  (tidak ada ketergantungan spasial pada galat)

$H_1: \lambda \neq 0$  (terdapat ketergantungan spasial pada galat)

Statistik ujinya adalah:

$$LM_{\lambda} = \frac{\left[ \frac{e'We}{\left(\frac{ee'}{n}\right)} \right]^2}{tr(W'W + WW)}$$

$H_0$  ditolak jika  $LM_{\lambda} > \chi^2_{\alpha(p)}$



# Regresi Spasial & Random Forest



Model yang akan digunakan dalam penelitian ini adalah:

## Model Regresi Klasik

$\rho = 0$  dan  $\lambda = 0$

Persamaan:

$$Y = X\beta + \epsilon$$

## Spatial Lag Model (SLM)

$\rho \neq 0$  dan  $\lambda = 0$

Persamaan:

$$y = \rho WY + X\beta + \epsilon$$
$$\epsilon \sim NIID(0, \sigma^2 I_n)$$

## Spatial Error Model (SEM)

$\rho = 0$  dan  $\lambda \neq 0$

Persamaan:

$$y = X\beta + u$$
$$u = \lambda \underline{W}u + \epsilon$$
$$\epsilon \sim NIID(0, \sigma^2 I_n)$$

## Random Forest

Random Forest bekerja dengan membagi data menjadi beberapa subset dan membangun pohon keputusan untuk setiap subset. Hasil dari pohon-pohon tersebut digabungkan untuk memberikan hasil prediksi yang lebih baik.





# Penentuan Model Terbaik

Pemilihan model terbaik dalam analisis regresi dilakukan dengan mempertimbangkan kriteria statistik tertentu untuk mengevaluasi kecocokan model

## Akaike Information Criterion (AIC)

Semakin kecil nilai AIC, semakin baik model tersebut.

$$AIC = -2 \ln(\hat{L}) + 2k$$

## Koefisien Determinasi (R<sup>2</sup>)

Semakin besar nilai R<sup>2</sup>, semakin baik model tersebut.

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

## Bayesian Information Criterion (BIC)

Semakin rendah nilai BIC, semakin baik model tersebut.

$$BIC = -2 \ln(\hat{L}) + k \ln(n)$$



# 04 HASIL DAN PEMBAHASAN





# Statistika Deskriptif



Berikut ini merupakan hasil statistika deskriptif dari dataset kami:

	Y	GiniRatio	RLS	MEDIS	MOTOR
count	38.000000	38.000000	38.000000	38.000000	3.800000e+01
mean	9.782368	0.331816	8.463947	48.157895	5.161917e+05
std	4.213447	0.037126	1.703252	34.714688	4.988880e+05
min	3.060000	0.233000	5.080000	11.000000	1.403630e+05
25%	6.517500	0.309250	7.477500	35.000000	2.515750e+05
50%	9.215000	0.329000	8.060000	40.500000	3.949030e+05
75%	12.042500	0.348500	9.832500	56.250000	5.550472e+05
max	20.830000	0.435000	12.110000	228.000000	3.034754e+06

Diperoleh terdapat 38 jumlah observasi yang mewakili kabupaten kota di Jawa Timur. Kemudian diperoleh nilai statistika deskriptif untuk masing-masing variabel yang kami gunakan. \*Y=Persentase Kemiskinan



# Hasil Uji Dependensi Spasial

Berikut disajikan diagnostic regression dari dataset kami terlebih dahulu:

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          30.824

TEST ON NORMALITY OF ERRORS
TEST            DF            VALUE            PROB
Jarque-Bera           2            0.713            0.7002

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST            DF            VALUE            PROB
Breusch-Pagan test     4            1.745            0.7825
Koenker-Bassett test   4            2.586            0.6293

[ ] 1 vif(lm(formula, data = df))

GiniRatio: 1.64422800376622 RLS: 1.65922928119909 MEDIS: 8.6168568374322 MOTOR: 8.75864003463505
```

Nilai VIF untuk X3 dan X4 > 5, bahkan mendekati 9, namun masih <10, menunjukkan indikasi multikolinearitas masih bisa ditoleransi antara X3 dan X4 dengan variabel lain. Sedangkan X1 dan X2 memiliki VIF < 5, artinya tidak bermasalah

Kesimpulan : Residual model berdistribusi normal dan memiliki variansi yang konstan (homoskedastis), tidak terjadi multiko.





# Hasil Uji Dependensi Spasial

Berikut disajikan hasil uji dependensi spasial dataset kami menggunakan Moran's I

Variabel Penelitian	Nama Variabel	Moran's I Statistic	p-value	Interpretasi
Variabel Dependen	Tingkat Kemiskinan (Y)	0.1998	0.03846	Signifikan
Variabel Independen	Gini Ratio (X1)	0.1817	0.05200	Hampir Signifikan
Variabel Independen	Rata-rata Lama Sekolah (X2)	0.2857	0.00474	Signifikan
Variabel Independen	Tenaga Medis (X3)	0.1337	0.04277	Signifikan
Variabel Independen	Pengguna Kendaraan Bermotor (X4)	0.1984	0.00664	Signifikan





# Hasil Uji Dependensi Spasial

Hasil Moran's I untuk masing-masing Model Regresi adalah sebagai berikut:

```
Moran I test under randomisation
```

```
data: model_ols$residuals  
weights: ww
```

```
Moran I statistic standard deviate = 1.1151, p-value = 0.1324  
alternative hypothesis: greater
```

```
sample estimates:
```

Moran I statistic	Expectation	Variance
0.09647716	-0.02702703	0.01226684





# Hasil Uji Dependensi Spasial

Hasil Moran's I untuk masing-masing Model Regresi adalah sebagai berikut:

```
Moran I test under randomisation
```

```
data: model_slm$residuals  
weights: ww
```

```
Moran I statistic standard deviate = 0.67997, p-value = 0.2483  
alternative hypothesis: greater
```

```
sample estimates:
```

Moran I statistic	Expectation	Variance
0.04823760	-0.02702703	0.01225178





# Hasil Uji Dependensi Spasial

Hasil Moran's I untuk masing-masing Model Regresi adalah sebagai berikut:

```
Moran I test under randomisation
```

```
data:  model_sem$residuals  
weights:  ww
```

```
Moran I statistic standard deviate = 0.19616, p-value = 0.4222  
alternative hypothesis: greater
```

```
sample estimates:
```

Moran I statistic	Expectation	Variance
-0.005281305	-0.027027027	0.012289860





# Hasil Uji Dependensi Spasial

Beberapa pendekatan lain selain Morans'I adalah sebagai berikut:

## DIAGNOSTICS FOR SPATIAL DEPENDENCE

- SARERR -

TEST	MI/DF	VALUE	PROB
Moran's I (error)	-0.0140	0.297	0.7665
Lagrange Multiplier (lag)	1	0.246	0.6199
Robust LM (lag)	1	0.279	0.5977
Lagrange Multiplier (error)	1	0.013	0.9086
Robust LM (error)	1	0.046	0.8308
Lagrange Multiplier (SARMA)	2	0.292	0.8643

Nilai p-value yang  $>0.05$  pada Uji Lagrange Multiplier Lag dan Error menyarankan bahwa **model OLS sudah cukup** dan tidak perlu beralih ke model SLM atau SEM.







# Hasil Analisis OLS, SLM, SEM

Berikut merupakan hasil pemodelan dari model OLS:

```
Call:
lm(formula = formula, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2188 -1.3762 -0.0764  1.1948  4.3407

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.263e+01  3.316e+00   9.840 2.42e-11 ***
GiniRatio   -3.163e+01  1.265e+01  -2.500  0.0176 *
RLS         -1.508e+00  2.770e-01  -5.445 4.96e-06 ***
MEDIS        6.931e-02  3.097e-02   2.238  0.0321 *
MOTOR       -5.665e-06  2.173e-06  -2.607  0.0136 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.228 on 33 degrees of freedom
Multiple R-squared:  0.7506,    Adjusted R-squared:  0.7203
F-statistic: 24.83 on 4 and 33 DF,  p-value: 1.5e-09
```

Model OLS ini adalah model yang baik dan signifikan secara statistik. Model ini mampu menjelaskan sekitar 72–75% variasi dalam variabel dependen. Semua variabel independen yang dimasukkan (GiniRatio, RLS, MEDIS, MOTOR) terbukti memiliki pengaruh yang signifikan terhadap variabel dependen. GiniRatio dan RLS memiliki pengaruh negatif yang kuat, MEDIS memiliki pengaruh positif, sedangkan MOTOR memiliki pengaruh negatif yang signifikan secara statistik namun sangat kecil secara praktis.

Persamaan:

$$Y = \beta_0 + \beta_1 \cdot \text{GiniRatio} + \beta_2 \cdot \text{RLS} + \beta_3 \cdot \text{MEDIS} + \beta_4 \cdot \text{MOTOR} + \varepsilon$$

Substitusi dari output:

$$Y = 32.63 - 31.63 \cdot \text{GiniRatio} - 1.508 \cdot \text{RLS} + 0.06931 \cdot \text{MEDIS} - 5.665 \times 10^{-6} \cdot \text{MOTOR} + \varepsilon$$



# Hasil Analisis OLS, SLM, SEM

```
Call:lagsarlm(formula = formula, data = df, listw = ww)

Residuals:
    Min       1Q   Median       3Q      Max
-4.29567 -1.30049  0.11694  1.16984  3.93045

Type: lag
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.5588e+01  3.4005e+00 10.4655 < 2.2e-16
GiniRatio    -3.3308e+01  1.1377e+01 -2.9277  0.003414
RLS          -1.6613e+00  2.7014e-01 -6.1498  7.757e-10
MEDIS         6.3486e-02  2.8242e-02  2.2480  0.024579
MOTOR        -5.0608e-06  2.0013e-06 -2.5288  0.011446

Rho: -0.031652, LR test value: 2.6049, p-value: 0.10653
Asymptotic standard error: 0.018098
      z-value: -1.7489, p-value: 0.080313
Wald statistic: 3.0586, p-value: 0.080313

Log likelihood: -80.38159 for lag model
ML residual variance (sigma squared): 4.0116, (sigma: 2.0029)
Number of observations: 38
Number of parameters estimated: 7
AIC: 174.76, (AIC for lm: 175.37)
LM test for residual autocorrelation
test value: 0.17527, p-value: 0.67547
```

Berikut merupakan hasil pemodelan dari Spatial Lag Model (SLM):

Karena p-value > 0.10, kita tidak menolak hipotesis nol. Ini berarti tidak ada lagi autokorelasi spasial yang signifikan dalam sisaan. Hal ini adalah pertanda baik, yang menunjukkan bahwa model SLM telah berhasil menangani dependensi spasial yang ada di dalam data.

Jika dilihat dari AIC, model SLM ini sedikit lebih baik daripada model regresi linear standar (OLS). Hal ini ditunjukkan oleh nilai AIC yang lebih rendah untuk model SLM (174.76) dibandingkan dengan AIC untuk model linear (175.37).

Persamaan:

$$Y = \rho WY + \beta_0 + \beta_1 \cdot \text{GiniRatio} + \beta_2 \cdot \text{RLS} + \beta_3 \cdot \text{MEDIS} + \beta_4 \cdot \text{MOTOR} + \varepsilon$$

Substitusi dari output:

$$Y = -0.03165 \cdot WY + 35.58 - 33.30 \cdot \text{GiniRatio} - 1.6613 \cdot \text{RLS} + 0.0635 \cdot \text{MEDIS} - 5.0608 \times 10^{-6} \cdot \text{MOTOR} + \varepsilon$$







# Hasil Analisis OLS, SLM, SEM

```
Call:errorsarlm(formula = formula, data = df, listw = ww)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.191796	-1.501315	-0.035765	1.132801	3.706161

Type: error

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.2422e+01	3.2303e+00	10.0370	< 2.2e-16
GiniRatio	-2.5515e+01	1.2571e+01	-2.0296	0.04239
RLS	-1.6707e+00	2.7546e-01	-6.0652	1.318e-09
MEDIS	5.3759e-02	2.9571e-02	1.8179	0.06907
MOTOR	-4.6618e-06	2.0835e-06	-2.2375	0.02525

Lambda: 0.075242, LR test value: 1.0575, p-value: 0.30379

Asymptotic standard error: 0.049041

z-value: 1.5342, p-value: 0.12497

Wald statistic: 2.3539, p-value: 0.12497

Log likelihood: -81.15528 for error model

ML residual variance (sigma squared): 4.098, (sigma: 2.0244)

Number of observations: 38

Number of parameters estimated: 7

AIC: 176.31, (AIC for lm: 175.37)

Berikut merupakan hasil pemodelan dari Spatial Error Model (SEM):

Persamaan:

$$Y = \beta_0 + \beta_1 \cdot \text{GiniRatio} + \beta_2 \cdot \text{RLS} + \beta_3 \cdot \text{MEDIS} + \beta_4 \cdot \text{MOTOR} + u$$

$$u = \lambda W u + \varepsilon$$

Substitusi dari output:

$$Y = 32.42 - 25.52 \cdot \text{GiniRatio} - 1.6707 \cdot \text{RLS} + 0.0538 \cdot \text{MEDIS} - 4.6618 \times 10^{-6} \cdot \text{MOTOR} + u$$

$$u = 0.0752 \cdot W u + \varepsilon$$

model SEM ini tidak memberikan perbaikan yang berarti karena koefisien spasial (Lambda) tidak signifikan dan nilai AIC model SEM lebih tinggi daripada AIC model linear. Dependensi spasial dalam galat bukanlah masalah yang signifikan dalam dataset ini.



# Random Forest Spasial

Model Random Forest spasial diestimasi menggunakan fungsi `rf_spatial()` pada data sebanyak 38 kabupaten/kota di Provinsi Jawa Timur. Model ini dilatih dengan 500 decision trees dan parameter `mtry` sebesar 2.

## Random forest parameters

- Type:	Regression
- Number of trees:	500
- Sample size:	38
- Number of predictors:	4
- Mtry:	2
- Minimum node size:	5







# Model Performance

Metrik	Nilai
R-squared (OOB)	0.6534121
R-squared (Prediksi)	0.9525424
Pseudo R-squared	0.9759828
RMSE (OOB)	2.480529
RMSE	1.116
Normalized RMSE	0.201991

- Model ini memiliki pseudo R-squared yang sangat tinggi (0.9759), menunjukkan bahwa hampir 98% variasi data kemiskinan dapat dijelaskan oleh model. Nilai RMSE sebesar 1.116 menunjukkan tingkat error prediksi yang rendah, jauh di bawah rata-rata kemiskinan di provinsi ini (~9.78%).
- Selain itu, hasil uji out-of-bag (OOB) juga mendukung akurasi model, dengan R-squared OOB sebesar 0.6534. Ini menandakan bahwa prediksi model cukup stabil, bahkan untuk data yang tidak pernah dilatih secara langsung. Secara keseluruhan, metrik ini menunjukkan bahwa model Random Forest Spatial sangat akurat dan bagus untuk tujuan prediksi.



# Variable Importance

Variable	Importance
X2	3.762
X1	1.448
X4	1.370
X3	0.794

- Menurut Spatial Random Forest, X2 / Rata Lama Sekolah merupakan variabel paling berpengaruh diantara yang lain





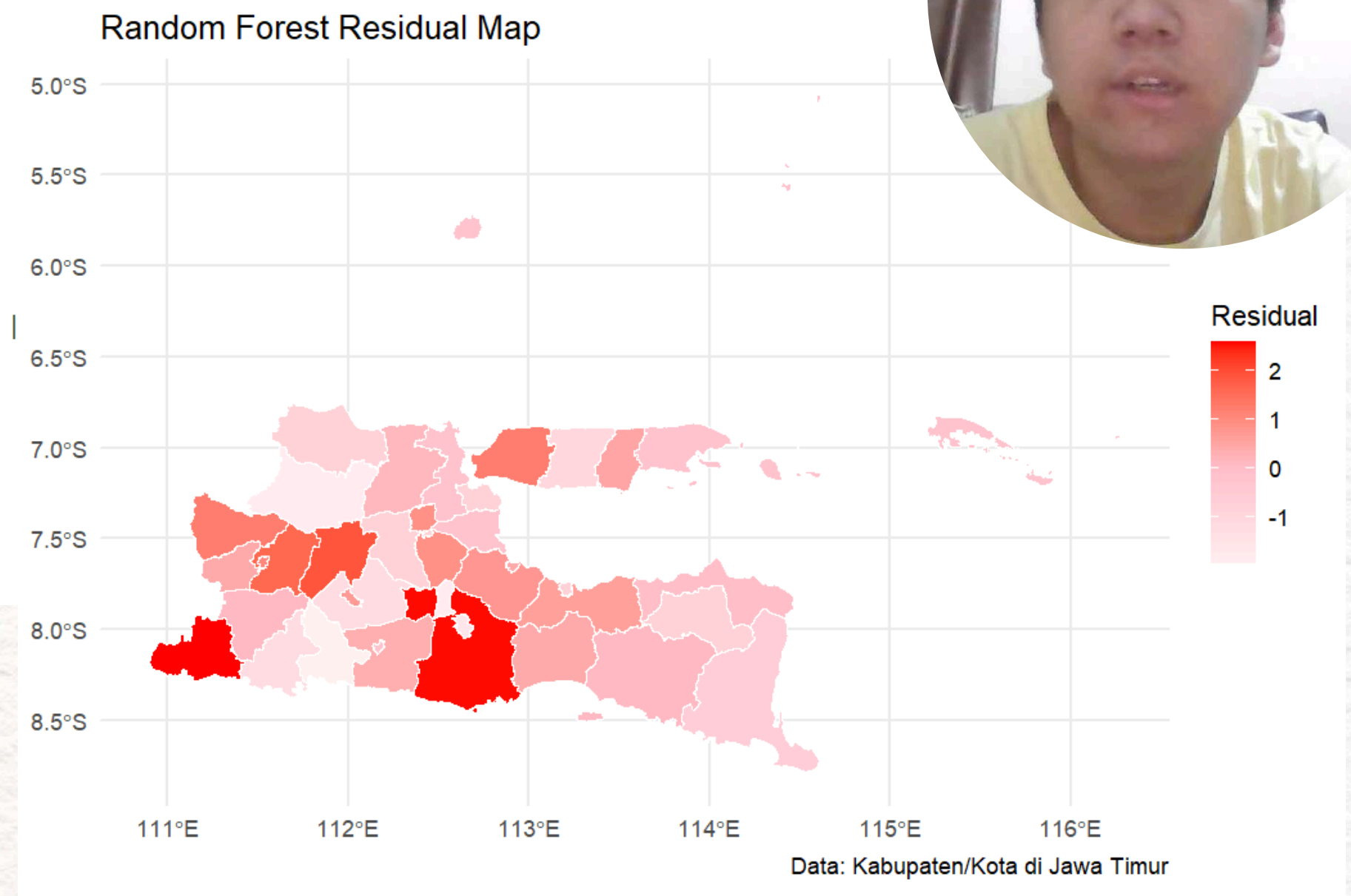
# Residual Analysis



- Normality:
  - Shapiro-Wilks W: 0.974
  - p-value : 0.5099
  - Interpretation : Residuals are normal

- Spatial autocorrelation:

Distance	Moran's I	P value	Interpretation
0.0	-0.037	0.900	No spatial correlation
1.0	-0.039	0.265	No spatial correlation
2.0	-0.033	0.592	No spatial correlation
3.0	-0.033	0.532	No spatial correlation



- Tidak ada spatial clustering di residual
- Model menangkap spatial heterogenitas secara baik

- Orange/Red: Positive residuals (under-prediction).
- Pink: Negative residuals (over-prediction).



# Key Points

- Model prediksi terbaik dibanding model regresi.
- Tidak ada bias spasial pada residual → Validasi kekuatan model.
- Faktor pendidikan ( $X_2$ ) tetap jadi faktor utama penurunan kemiskinan.
- Cocok untuk perencanaan kebijakan & intervensi yang tepat sasaran.







# Perbandingan Model

- OLS memberikan hasil R-squared sekitar 0.75. Ini menunjukkan bahwa sekitar 75% variasi tingkat kemiskinan dapat dijelaskan oleh model.
  - SAR dan SEM menunjukkan hasil yang mirip dengan OLS, tetapi koefisien spasial ( $\rho$  untuk SAR,  $\lambda$  untuk SEM) ternyata tidak signifikan ( $p\text{-value} > 0.05$ ).
  - Meskipun demikian, SEM memiliki nilai AIC dan BIC yang sedikit lebih rendah daripada OLS, yang berarti modelnya secara statistik sedikit lebih efisien. Namun, perbedaan tersebut kecil dan tidak signifikan secara substansial.
- Secara keseluruhan OLS, SAR, SEM cocok untuk interpretasi arah & pengaruh variabel.
  - Spatial Random Forest unggul dalam prediksi, mampu menangkap hubungan non-linear dan spasial yang kompleks.
  - Residual dari semua model tidak menunjukkan pola spasial yang signifikan, tetapi SRF tetap menjadi model prediksi yang paling akurat.
  - Untuk analisis kebijakan, OLS cukup. Namun, untuk prediksi dan perencanaan berbasis data, SRF direkomendasikan.



# 05 KESIMPULAN





# Kesimpulan



## 1. Model Regresi Spasial

- OLS, SAR, dan SEM menunjukkan R-squared sekitar 0.75.
- Koefisien spasial SAR dan SEM tidak signifikan → menunjukkan efek spasial global tidak dominan.
- OLS cukup memadai untuk interpretasi arah pengaruh antar-variabel

## 2. Model Spatial Random Forest (SRF)

- SRF memiliki pseudo R-squared tertinggi (0.9759) dan RMSE terendah (1.116).
- Residual SRF terdistribusi normal dan tidak ada autokorelasi spasial.
- Faktor yang paling berpengaruh: Rata-rata Lama Sekolah (X2) → menunjukkan pentingnya pendidikan dalam menurunkan kemiskinan.

## 3. Implikasi

- Model regresi spasial cocok untuk interpretasi hubungan.
- SRF menjadi model terbaik untuk prediksi kemiskinan yang akurat.
- Kebijakan pengentasan kemiskinan sebaiknya fokus pada peningkatan akses dan kualitas pendidikan.



# Daftar Pustaka



A.A. Maulana, A. Fauzan. "Spatially informed insights: Modeling percentage poverty in East Java Province using SEM with spatial weight variations," BAREKENG: Journal of Mathematics and Its Applications, vol. 18, no. 6, pp. 1317–1332, June 2024.

Badan Pusat Statistik (BPS), [Online]. Available: <https://www.bps.go.id/id> [Accessed: 26 May 2025].

Badan Pusat Statistik (BPS), "Persentase Penduduk Miskin di Jawa Timur September 2024 turun menjadi 9,56 persen", [Online]. Available: <https://jatim.bps.go.id/id/pressrelease/2025/01/15/1474/persentase-penduduk-miskin-di-jawa-timur-september-2024-turun-menjadi-9-56-persen.html> [Accessed: 5 June 2025].

E.W. Fox, J.M.V. Hoef. "Comparing spatial regression to random forests for large environmental data sets," PLoS ONE, vol. 15, no. 3, March 2020.

IBM, "Apa itu random forest?", [Online]. Available: <https://www.ibm.com/id-id/think/topics/random-forest> [Accessed: 7 June 2025].

L. Anselin, Spatial econometrics: methods and models. New York: McGraw Hill, 1953.

N.U. Ayudia, D.E. Putri. "Pemodelan Spatial Autoregressive Model (SAR) pada kasus kemiskinan di Jawa Timur," Jurnal Gaussian, vol. 13, no. 2, pp. 308–1318, November 2024.

R. Hasanah, Syaparuddin, Rosmeli. "Pengaruh angka harapan hidup, rata-rata lama sekolah dan pengeluaran perkapita terhadap tingkat kemiskinan pada Kabupaten /Kota di Provinsi Jambi," E-Jurnal Perspektif Ekonomi Dan Pembangunan Daerah, vol. 10, no. 3, pp. 223–232, September 2021.



# **Link Video Presentasi**

**<https://youtu.be/ODEhQMngCi8>**





# TERIMA KASIH

---

**Mohon maaf atas kesalahan dan  
kekurangan selama presentasi berlangsung.**