



UNIVERSITAS INDONESIA

**ANALISIS REGRESI SPASIAL (SAR/SEM) PADA SEKTOR
PEKERJAAN KESEHATAN DAN SOSIAL DI INGGRIS**

LAPORAN

**AMMAR HANAFI (2206051582)
NORMAN MOWLANA AZIZ (2206025470)
KIRONO DWI SAPUTRO (2106656365)
DEVANA SOLEA (2306262402)**

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

Daftar Isi

BAB I PENDAHULUAN	4
1.1 Latar Belakang	4
1.2 Rumusan Masalah	5
1.3 Tujuan	5
BAB II LANDASAN TEORI	6
2.1 Statistika Deskriptif	6
2.1.1 Ukuran Pemusatan	6
2.1.2 Ukuran Penyebaran	6
2.2 Regresi Linear Berganda (Ordinary Least Squares)	7
2.3 Matriks Pembobot Spasial (Spatial Weights Matrix)	7
2.3.1 Rook Contiguity	7
2.3.2 Queen Contiguity	8
2.3.3 Bishop Contiguity	8
2.4 Indeks Moran (Moran's I)	8
2.5 Uji Lagrange Multiplier (LM)	9
2.6 Spatial Lag Model (SAR)	9
2.7 Spatial Error Model (SEM)	10
BAB III METODE PENELITIAN	11
3.1 Data dan Variabel	11
3.2 Tahapan Analisis	11
BAB IV ANALISIS DAN PEMBAHASAN	12
4.1 Analisis Deskriptif dan Eksplorasi Data (EDA)	12
4.1.1 Statistik Deskriptif	12
4.1.2 Visualisasi Data	12
4.1.3 Peta Sebaran	14
4.2 Hasil Regresi OLS (Global)	15
4.3 Uji Diagnostik Regresi Global	16
4.4 Diagnostik Dependensi Spasial	16
4.5 Hasil Regresi Spasial (SEM)	17
4.6 Perbandingan Model	19
4.7 Interpretasi	20
BAB V KESIMPULAN	21
DAFTAR PUSTAKA	24
LAMPIRAN	25

Daftar Tabel

1	Sampel Data (5 Baris Pertama)	11
2	Statistik Deskriptif Variabel Penelitian	12
3	Ringkasan Hasil Regresi OLS	15
4	Hasil Uji Diagnostik OLS	16
5	Variance Inflation Factor (VIF)	16
6	Hasil Uji Diagnostik Spasial	16
7	Ringkasan Hasil Spatial Error Model (SEM)	17
8	Perbandingan Kebaikan Model	19

Daftar Gambar

1	Histogram Distribusi Variabel	13
2	Boxplot Variabel (Deteksi Outlier)	13
3	Matriks Korelasi Antar Variabel	14
4	Peta Sebaran Tenaga Kerja Sektor Kesehatan (Quantile)	15
5	Peta Residual: OLS vs SAR vs SEM	17
6	Peta Residual Model SEM	19
7	Peta Kluster LISA (Local Indicators of Spatial Association)	21

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor kesehatan dan pekerjaan sosial merupakan salah satu sektor strategis yang memiliki peran fundamental dalam mendukung kualitas hidup masyarakat serta keberlanjutan pembangunan ekonomi suatu wilayah. Keberadaan tenaga kerja pada sektor ini tidak hanya berkontribusi terhadap peningkatan derajat kesehatan masyarakat, tetapi juga berperan dalam memperkuat sistem perlindungan sosial dan stabilitas sosial ekonomi secara keseluruhan. Dalam konteks pembangunan wilayah, distribusi tenaga kerja pada sektor *Human Health and Social Work* sering kali menunjukkan variasi yang cukup besar antar daerah. Variasi tersebut dipengaruhi oleh berbagai faktor struktural, seperti tingkat pendidikan, perkembangan sektor keuangan dan asuransi, sektor informasi dan komunikasi, serta aktivitas real estat.

Secara teoritis, wilayah dengan tingkat pendidikan yang lebih tinggi cenderung memiliki kapasitas sumber daya manusia yang lebih baik untuk mendukung sektor kesehatan dan layanan sosial. Pendidikan berperan dalam meningkatkan kompetensi tenaga kerja, memperluas akses terhadap informasi, serta mendorong inovasi dalam penyediaan layanan publik. Selain itu, sektor keuangan dan asuransi berpotensi mempengaruhi pembiayaan layanan kesehatan, sementara sektor real estat mencerminkan dinamika pembangunan wilayah yang dapat berdampak pada kebutuhan dan penyediaan fasilitas kesehatan. Dengan demikian, hubungan antar sektor ekonomi tersebut bersifat kompleks dan saling terkait.

Namun demikian, dalam analisis data kewilayahan, terdapat karakteristik khusus yang membedakannya dari data non-spasial, yaitu adanya kemungkinan ketergantungan spasial (*spatial dependence*). Ketergantungan spasial mengacu pada kondisi di mana nilai suatu variabel pada suatu wilayah tidak berdiri sendiri, melainkan dipengaruhi oleh nilai variabel pada wilayah sekitarnya. Fenomena ini sejalan dengan Hukum Geografi Pertama Tobler yang menyatakan bahwa “segala sesuatu saling berhubungan, tetapi hal-hal yang berdekatan memiliki hubungan yang lebih kuat dibandingkan yang berjauhan”. Dalam konteks tenaga kerja sektor kesehatan, wilayah yang berdekatan secara geografis sering kali memiliki karakteristik ekonomi dan sosial yang serupa, sehingga memungkinkan terbentuknya pola pengelompokan (*clustering*).

Apabila ketergantungan spasial diabaikan, penggunaan model regresi linear klasik atau *Ordinary Least Squares* (OLS) berpotensi menghasilkan estimator yang tidak efisien dan bias dalam inferensi statistik. Salah satu asumsi penting dalam OLS adalah independensi error, yaitu tidak adanya korelasi antar residual observasi. Jika terdapat autokorelasi spasial pada residual, maka varians estimator menjadi tidak akurat sehingga uji signifikansi parameter dapat menyesatkan. Oleh karena itu, diperlukan pendekatan pemodelan yang secara eksplisit mengakomodasi struktur spasial dalam data.

Regresi spasial merupakan pengembangan dari model regresi klasik yang dirancang untuk menangani adanya interaksi atau dependensi antar wilayah. Dua model yang umum digunakan adalah *Spatial Autoregressive Model* (SAR), yang memasukkan pengaruh variabel dependen dari wilayah tetangga ke dalam model, serta *Spatial Error*

Model (SEM), yang mengakomodasi ketergantungan spasial pada komponen error. Pemilihan model yang tepat sangat penting untuk memastikan bahwa hubungan antar variabel ekonomi yang dianalisis dapat diinterpretasikan secara valid dan konsisten.

Dengan mempertimbangkan pentingnya sektor kesehatan dan pekerjaan sosial dalam pembangunan wilayah serta potensi adanya pola spasial dalam distribusinya, penelitian ini menggunakan pendekatan regresi spasial untuk mengidentifikasi faktor-faktor yang mempengaruhi variasi tenaga kerja pada sektor tersebut. Pendekatan ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai dinamika ekonomi kewilayahan dan menjadi dasar dalam perumusan kebijakan pembangunan yang lebih tepat sasaran.

1.2 Rumusan Masalah

1. Apakah terdapat ketergantungan spasial (spatial dependence) dalam model proporsi pekerja sektor kesehatan dan sosial di Inggris?
2. Bagaimana pengaruh variabel sektor pekerjaan (Education, Finance, Info-Comm, Real Estate) terhadap tingkat pekerja di sektor Kesehatan dan Sosial (Human Health and Social Work)?
3. Model spasial manakah (SAR atau SEM) yang terbaik untuk menjelaskan data tersebut?

1.3 Tujuan

Tujuan dari analisis ini adalah untuk mengidentifikasi dan memodelkan hubungan antara variabel-variabel prediktor dengan variabel respon dengan memperhitungkan efek spasial, serta menentukan model terbaik antara OLS, SAR, dan SEM.

BAB II LANDASAN TEORI

2.1 Statistika Deskriptif

Statistika deskriptif merupakan cabang ilmu statistika yang digunakan untuk menggambarkan karakteristik data tanpa melakukan generalisasi terhadap populasi yang lebih luas. Dalam penelitian spasial, statistika deskriptif tidak hanya menjelaskan distribusi numerik tetapi juga pola geografis antar wilayah.

2.1.1 Ukuran Pemusatan

Mean Mean atau rata-rata dihitung dengan rumus:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

di mana x_i adalah nilai pengamatan ke- i dan n adalah jumlah observasi.

Median Median adalah nilai tengah setelah data diurutkan.

Kuartil Kuartil membagi data menjadi empat bagian:

- Q1 (25%)
- Q2 (Median)
- Q3 (75%)

2.1.2 Ukuran Penyebaran

Varians

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2)$$

Standar Deviasi

$$s = \sqrt{s^2} \quad (3)$$

Koefisien Variasi

$$CV = \frac{s}{\bar{x}} \times 100\% \quad (4)$$

Koefisien variasi digunakan untuk membandingkan tingkat variasi antar variabel dengan skala yang berbeda.

2.2 Regresi Linear Berganda (Ordinary Least Squares)

Regresi linear berganda digunakan untuk menganalisis pengaruh beberapa variabel independen terhadap satu variabel dependen.

Model umum regresi linear berganda adalah:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \quad (5)$$

dengan:

- Y_i = variabel dependen
- X_{ki} = variabel independen ke- k
- β_k = parameter regresi
- ε_i = error

2.3 Matriks Pembobot Spasial (Spatial Weights Matrix)

Matriks pembobot spasial W adalah komponen kunci dalam analisis spasial yang merepresentasikan struktur hubungan antar lokasi. Dalam analisis ini, digunakan pembobot **Queen Contiguity**, di mana lokasi i dan j dianggap bertetangga jika mereka berbagi sisi atau sudut perbatasan yang sama.

Matriks pembobot spasial (W) digunakan untuk merepresentasikan hubungan kedekatan antar wilayah. Elemen matriks w_{ij} menunjukkan tingkat keterkaitan antara wilayah i dan wilayah j .

Secara umum, matriks pembobot spasial didefinisikan sebagai:

$$W = [w_{ij}] \quad (6)$$

dengan:

$$w_{ij} = \begin{cases} 1, & \text{jika wilayah } i \text{ bertetangga dengan wilayah } j \\ 0, & \text{lainnya} \end{cases} \quad (7)$$

Untuk menjaga konsistensi skala, dilakukan normalisasi baris:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}} \quad (8)$$

Sehingga setiap baris memiliki jumlah bobot sebesar 1.

2.3.1 Rook Contiguity

Rook contiguity mendefinisikan dua wilayah sebagai bertetangga apabila keduanya memiliki batas sisi (edge) yang sama. Dengan kata lain, hubungan spasial hanya terjadi apabila terdapat perpotongan garis batas wilayah.

Secara matematis:

$$w_{ij}^{rook} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi sisi} \\ 0, & \text{lainnya} \end{cases} \quad (9)$$

Metode ini cenderung menghasilkan jumlah tetangga yang lebih sedikit dibandingkan Queen contiguity.

2.3.2 Queen Contiguity

Queen contiguity mendefinisikan dua wilayah sebagai bertetangga apabila keduanya memiliki batas sisi atau titik sudut (vertex) yang sama. Konsep ini menyerupai pergerakan bidak ratu dalam permainan catur.

$$w_{ij}^{queen} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi sisi atau titik sudut} \\ 0, & \text{lainnya} \end{cases} \quad (10)$$

Queen contiguity biasanya menghasilkan matriks dengan konektivitas lebih tinggi dibandingkan rook.

2.3.3 Bishop Contiguity

Bishop contiguity mendefinisikan dua wilayah sebagai bertetangga hanya apabila keduanya berbagi titik sudut (vertex) tetapi tidak berbagi sisi. Konsep ini menyerupai pergerakan bidak gajah dalam permainan catur.

$$w_{ij}^{bishop} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi titik sudut saja} \\ 0, & \text{lainnya} \end{cases} \quad (11)$$

Model bishop menghasilkan struktur ketetanggaan yang lebih selektif dibandingkan rook maupun queen.

2.4 Indeks Moran (Moran's I)

Indeks Moran digunakan untuk mengukur autokorelasi spasial global.

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (12)$$

dengan:

- w_{ij} = bobot spasial
- $S_0 = \sum_i \sum_j w_{ij}$

Interpretasi:

- $I > 0$ menunjukkan clustering
- $I < 0$ menunjukkan dispersi
- $I \approx 0$ menunjukkan pola acak

2.5 Uji Lagrange Multiplier (LM)

Uji Lagrange Multiplier (LM) digunakan untuk memilih antara model SAR dan SEM. Uji ini terdiri dari:

- **LM-Lag:** Menguji signifikansi lag spasial pada variabel dependen.
- **LM-Error:** Menguji signifikansi korelasi spasial pada error.

Jika LM-Lag signifikan dan LM-Error tidak, maka model SAR dipilih (dan sebaliknya). Jika keduanya signifikan, maka dilihat varian Robust LM.

2.6 Spatial Lag Model (SAR)

Model *Spatial Autoregressive* (SAR) digunakan ketika terdapat indikasi bahwa nilai variabel dependen pada suatu wilayah dipengaruhi secara langsung oleh nilai variabel dependen pada wilayah-wilayah tetangganya. Model ini merepresentasikan adanya efek interaksi atau *spillover effect*, di mana perubahan pada satu wilayah dapat menyebar ke wilayah lain melalui struktur kedekatan spasial yang didefinisikan dalam matriks pembobot W .

Secara matematis, model SAR dinyatakan sebagai:

$$y = \rho W y + X\beta + \varepsilon \quad (13)$$

di mana:

- y : vektor ($n \times 1$) variabel dependen.
- W : matriks pembobot spasial ($n \times n$) yang menggambarkan struktur kedekatan antar wilayah.
- ρ : parameter lag spasial yang mengukur kekuatan pengaruh spasial pada variabel dependen.
- $W y$: lag spasial dari variabel dependen.
- X : matriks ($n \times k$) variabel independen.
- β : vektor koefisien regresi ($k \times 1$).
- ε : vektor error yang diasumsikan berdistribusi normal dengan

$$\varepsilon \sim N(0, \sigma^2 I)$$

Jika persamaan (1) diturunkan lebih lanjut, maka dapat ditulis ulang sebagai:

$$(I - \rho W)y = X\beta + \varepsilon \quad (14)$$

Sehingga:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon \quad (15)$$

di mana I adalah matriks identitas ($n \times n$).

Persamaan (3) menunjukkan bahwa perubahan pada satu wilayah akan mempengaruhi wilayah lain melalui proses propagasi spasial. Oleh karena itu, dalam model SAR dikenal adanya:

- **Direct Effect** (efek langsung)
- **Indirect Effect** (efek tidak langsung / spillover)
- **Total Effect**

Model SAR digunakan ketika autokorelasi spasial terjadi pada variabel dependen dan terbukti signifikan melalui uji Moran's I atau uji Lagrange Multiplier (LM-Lag).

2.7 Spatial Error Model (SEM)

Model *Spatial Error Model* (SEM) digunakan ketika ketergantungan spasial tidak terjadi secara langsung pada variabel dependen, tetapi muncul pada komponen error. Ketergantungan ini biasanya disebabkan oleh adanya variabel yang tidak teramati (*omitted variables*) yang memiliki pola spasial.

Model SEM dinyatakan sebagai:

$$y = X\beta + u \quad (16)$$

dengan struktur error spasial:

$$u = \lambda W u + \varepsilon \quad (17)$$

di mana:

- y : vektor ($n \times 1$) variabel dependen.
- X : matriks variabel independen.
- β : vektor koefisien regresi.
- u : komponen error yang memiliki struktur spasial.
- W : matriks pembobot spasial.
- λ : parameter error spasial yang mengukur tingkat ketergantungan spasial pada residual.
- ε : error acak dengan asumsi:

$$\varepsilon \sim N(0, \sigma^2 I)$$

Jika persamaan (5) disubstitusikan ke dalam persamaan (4), maka diperoleh:

$$y = X\beta + (I - \lambda W)^{-1} \varepsilon \quad (18)$$

Model SEM menunjukkan bahwa korelasi spasial terjadi pada error, bukan pada variabel dependen secara langsung. Berbeda dengan SAR, model SEM tidak menghasilkan efek spillover langsung, tetapi memperbaiki bias yang disebabkan oleh struktur spasial pada residual.

BAB III

METODE PENELITIAN

3.1 Data dan Variabel

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

- **Variabel Dependen** (*health_social*): Jumlah populasi bekerja (16-74 tahun) yang bekerja di bidang kesehatan (*Human_health_and_social_worker*).
- **Variabel Independen** (*education*): Jumlah populasi bekerja (16-74 tahun) yang bekerja di bidang pendidikan (*Education*).
- **Variabel Independen** (*financial*): Jumlah populasi bekerja (16-74 tahun) yang bekerja di bidang finansial dan asuransi (*Financial_and_insurance*).
- **Variabel Independen** (*information*): Jumlah populasi bekerja (16-74 tahun) yang bekerja di bidang informasi dan komunikasi (*Information_and_communication*).
- **Variabel Independen** (*estate*): Jumlah populasi bekerja (16-74 tahun) yang bekerja di bidang real estat (*Real_Estate*).

Tabel 1 berikut menampilkan 5 baris pertama dari dataset yang digunakan.

Tabel 1: Sampel Data (5 Baris Pertama)

Region	Health & Social	Education	Financial	Info & Comm	Real Estate
Hartlepool	5829	3641	789	645	588
Middlesbrough	9116	5362	1358	1266	702
Redcar & Cleveland	8787	5545	1103	963	898
Stockton-on-Tees	12928	8721	2713	2086	1125
Darlington	7259	4211	1880	2165	500

3.2 Tahapan Analisis

1. **Eksplorasi Data:** Memeriksa struktur data dan variabel.
2. **Estimasi Model OLS:** Sebagai model dasar (baseline).
3. **Diagnostik Spasial:** Melakukan uji Moran's I pada residual OLS dan uji LM (Lag & Error) untuk mendeteksi autokorelasi spasial dan menentukan model yang tepat.
4. **Estimasi Model Spasial:** Mengestimasi model terpilih (SAR/SEM).
5. **Evaluasi Model:** Mengevaluasi model terbaik.
6. **Interpretasi dan Visualisasi Hasil:** Mencari kesimpulan dari hasil yang diperoleh

BAB IV ANALISIS DAN PEMBAHASAN

4.1 Analisis Deskriptif dan Eksplorasi Data (EDA)

Sebelum melakukan pemodelan, dilakukan pemeriksaan karakteristik data melalui statistik deskriptif dan visualisasi.

4.1.1 Statistik Deskriptif

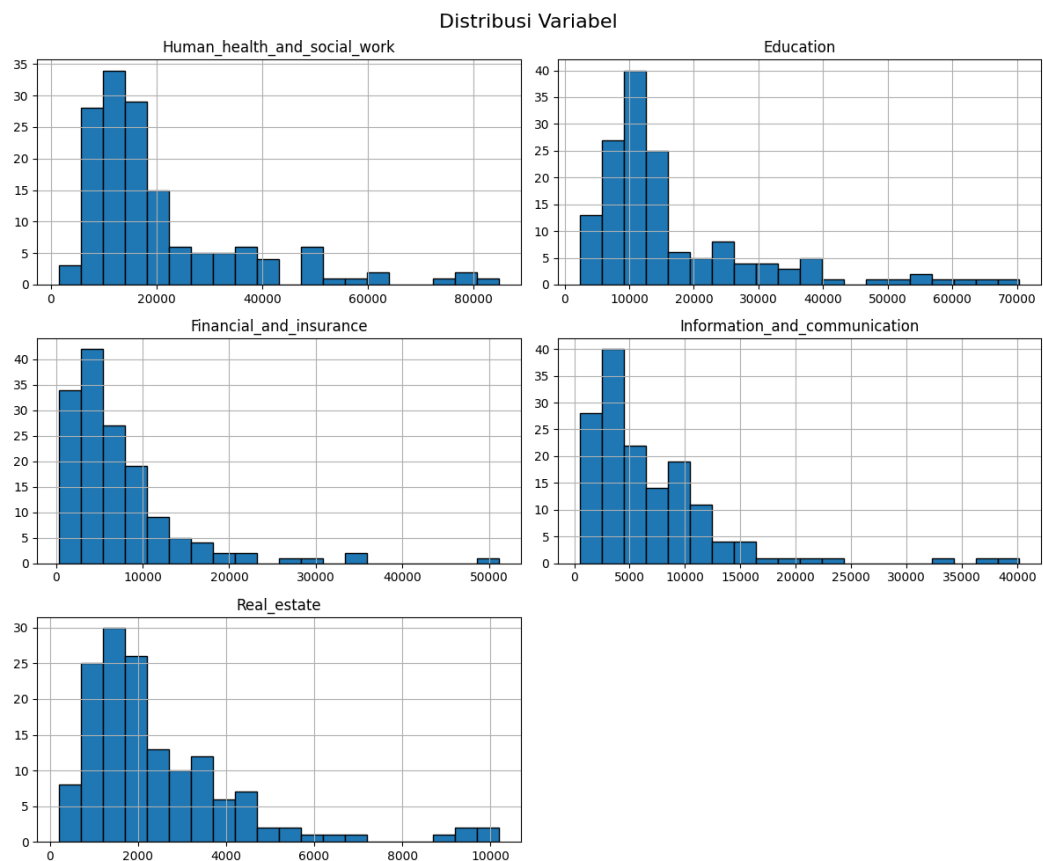
Tabel 2 menyajikan ringkasan statistik untuk variabel-variabel yang digunakan.

Tabel 2: Statistik Deskriptif Variabel Penelitian

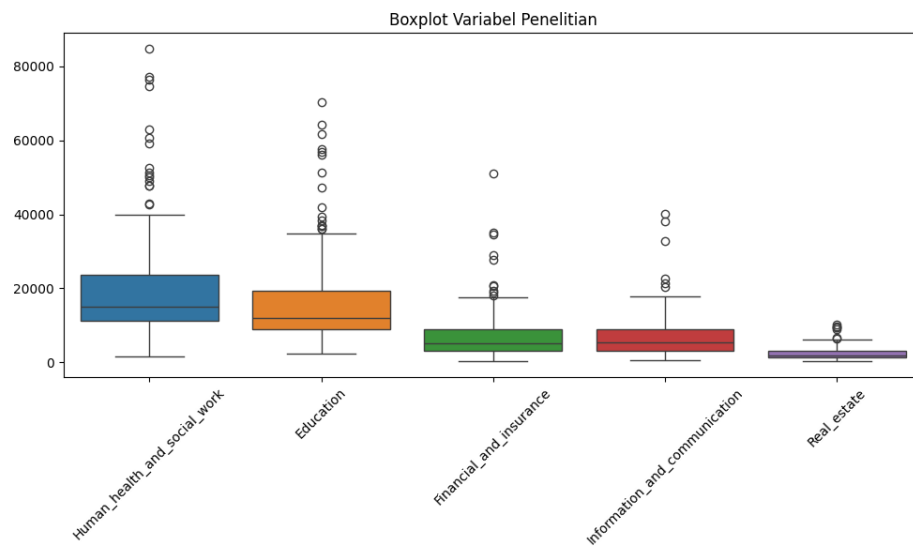
Variable	Count	Mean	Std	Min	25%	50%	75%	Max	Skew	Kurtosis
Health & Social	149	20947.9	15772.2	1585	11219	15118	23558	84796	1.92	3.72
Education	149	16712.7	13157.9	2386	8827	11951	19418	70363	2.00	3.99
Financial	149	7408.4	7010.5	387	3196	5190	9028	51117	2.97	12.50
Info & Comm	149	6874.8	6152.1	518	3018	5315	8995	40213	2.79	10.96
Real Estate	149	2466.2	1842.9	216	1286	1906	3191	10188	2.16	5.69

4.1.2 Visualisasi Data

Distribusi data dan hubungan antar variabel dapat dilihat pada gambar-gambar berikut.

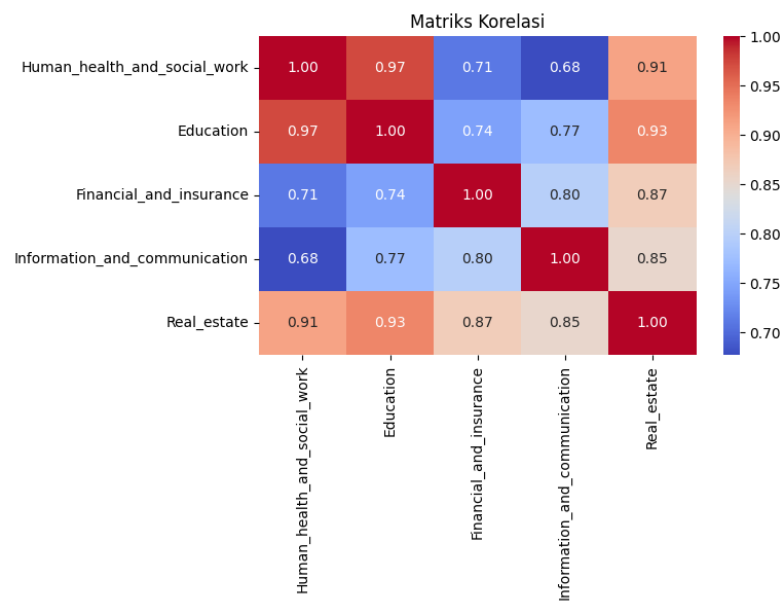


Gambar 1: Histogram Distribusi Variabel



Gambar 2: Boxplot Variabel (Deteksi Outlier)

Gambar 1 dan 2 menunjukkan bahwa sebagian besar variabel memiliki distribusi menjulur ke kanan (skewed right) dan terdapat beberapa pencilan (outlier), terutama pada sektor finansial.

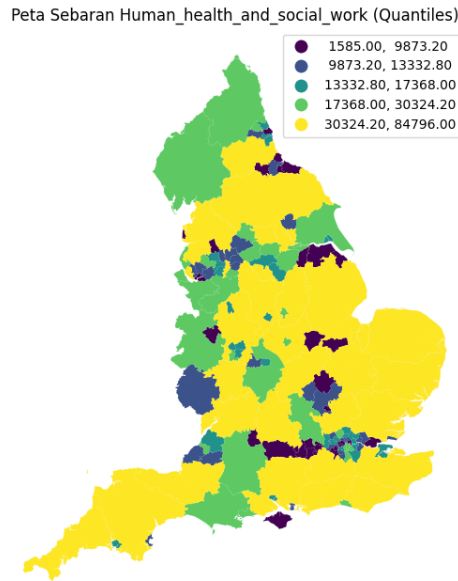


Gambar 3: Matriks Korelasi Antar Variabel

Matriks korelasi (Gambar 3) menunjukkan adanya hubungan linear yang cukup kuat antar variabel prediktor, yang perlu diwaspadai sebagai potensi multikolinearitas, namun VIF (Variance Inflation Factor) pada OLS perlu diperiksa lebih lanjut.

4.1.3 Peta Sebaran

Peta sebaran variabel respon (Tenaga Kerja Kesehatan & Sosial) ditampilkan dalam Gambar 4.



Gambar 4: Peta Sebaran Tenaga Kerja Sektor Kesehatan (Quantile)

4.2 Hasil Regresi OLS (Global)

Sebagai langkah awal, dilakukan estimasi menggunakan Ordinary Least Squares (OLS). Model untuk OLS global ini adalah sebagai berikut:

$$health_social = \beta_0 + \beta_1 education + \beta_2 financial + \beta_3 information + \beta_4 estate + \varepsilon \quad (19)$$

Hasil dari model regresi OLS global ini adalah sebagai berikut:

Tabel 3: Ringkasan Hasil Regresi OLS

Variabel	Koefisien	Std. Error	t-Statistic	Prob.
CONSTANT	1289.143	393.284	3.278	0.001
Education	1.066	0.054	19.912	0.000
Financial	-0.053	0.074	-0.714	0.476
Info & Comm	-0.725	0.075	-9.705	0.000
Real Estate	2.929	0.547	5.356	0.000

Model OLS menghasilkan nilai R^2 yang sangat tinggi yaitu **0.968**, yang menunjukkan model ini sangat baik dalam menjelaskan variasi data. Variabel **Education**, **Information_and_communication**, dan **Real_estate** berpengaruh signifikan terhadap variabel respon. Namun, diagnosis spasial diperlukan untuk memastikan validitas model.

4.3 Uji Diagnostik Regresi Global

Selanjutnya dilakukan uji diagnostik terhadap model regresi OLS yang dihasilkan.

Tabel 4: Hasil Uji Diagnostik OLS

Uji	Nilai Statistik	Probabilitas (p-value)
Shapiro-Wilk	0.9445	0.0000
Breush-Pagan	52.0116	0.0000

Berdasarkan Tabel 2, ditemukan bahwa residual pada model OLS global yang dihasilkan tidak berdistribusi normal dan bersifat heteroskedastik. Heteroskedastisitas dapat dilihat dari nilai VIF sebagai berikut:

Tabel 5: Variance Inflation Factor (VIF)

Variabel	VIF
const	2.8492
Education	9.0761
Financial_and_insurance	4.8988
Information_and_communication	3.8679
Real_estate	18.5849

4.4 Diagnostik Dependensi Spasial

Untuk mendeteksi adanya autokorelasi spasial, dilakukan uji pada residual OLS.

Tabel 6: Hasil Uji Diagnostik Spasial

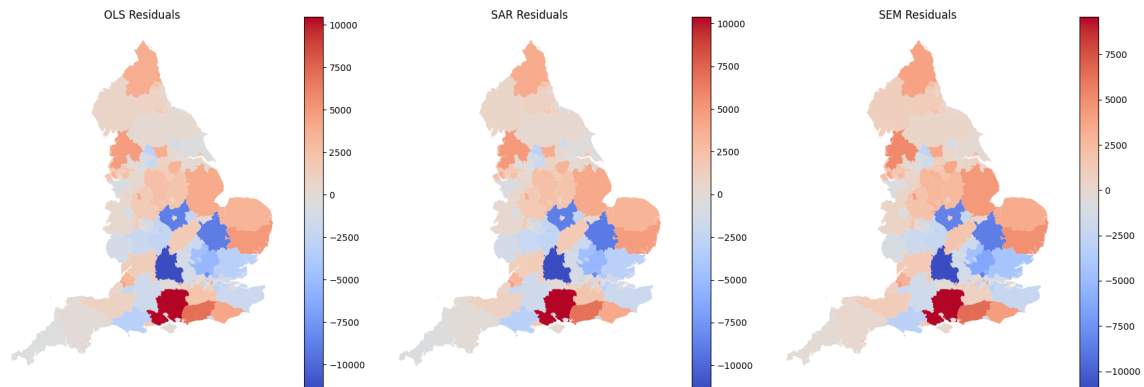
Uji	Nilai Statistik	Probabilitas (p-value)
Moran's I (error)	0.219	0.001
LM (Lag)	0.176	0.6746
Robust LM (Lag)	0.029	0.8647
LM (Error)	13.867	0.0002
Robust LM (Error)	13.719	0.0002

Berdasarkan Tabel 6:

- Uji **Moran's I** pada error signifikan ($p < 0.05$), menunjukkan adanya autokorelasi spasial positif pada residual.
- Uji **LM-Error** signifikan ($p = 0.0002$), sedangkan **LM-Lag** tidak signifikan ($p = 0.6746$).

Sesuai dengan kaidah keputusan Anselin, karena LM-Error signifikan dan LM-Lag tidak, maka model yang paling tepat adalah **Spatial Error Model (SEM)**. Ini mengindikasikan bahwa ketergantungan spasial terjadi melalui error, mungkin karena adanya

variabel yang tidak dimasukkan ke dalam model (omitted variables) yang memiliki pola spasial.



Gambar 5: Peta Residual: OLS vs SAR vs SEM

4.5 Hasil Regresi Spasial (SEM)

Berdasarkan hasil diagnostik, model SEM diestimasi menggunakan metode Maximum Likelihood.

Tabel 7: Ringkasan Hasil Spatial Error Model (SEM)

Variabel	Koefisien	Std. Error	z-Statistic	Prob.
CONSTANT	1096.602	395.642	2.772	0.006
Education	1.028	0.056	18.486	0.000
Financial	-0.025	0.073	-0.342	0.733
Info & Comm	-0.696	0.078	-8.866	0.000
Real Estate	3.095	0.554	5.587	0.000
Lambda (λ)	0.283	0.103	2.752	0.006

Berdasarkan hasil estimasi *Spatial Error Model* (SEM) pada Tabel di atas, diperoleh beberapa temuan penting terkait pengaruh variabel independen terhadap sektor *Human Health and Social Work* serta keberadaan dependensi spasial pada komponen error.

Pertama, nilai konstanta (CONSTANT) sebesar 1096.602 dengan nilai probabilitas 0.006 menunjukkan bahwa parameter intercept signifikan secara statistik pada tingkat signifikansi 1%. Hal ini mengindikasikan bahwa ketika seluruh variabel independen bernilai nol, terdapat estimasi dasar jumlah tenaga kerja kesehatan sebesar 1096 unit.

Variabel *Education* memiliki koefisien sebesar 1.0284 dan signifikan pada tingkat 1% (p-value = 0.000). Nilai z-statistic sebesar 18.486 menunjukkan bahwa pengaruh variabel ini sangat kuat secara statistik. Koefisien positif mengindikasikan bahwa peningkatan satu satuan pada sektor pendidikan akan meningkatkan sektor *Human Health and Social Work* sebesar 1.0284 satuan, dengan asumsi variabel lain konstan. Secara substantif, hasil ini menunjukkan bahwa wilayah dengan tingkat aktivitas pendidikan yang lebih tinggi cenderung memiliki tenaga kerja sektor kesehatan dan pekerjaan sosial yang lebih besar. Temuan ini konsisten dengan teori modal manusia (*human capital*

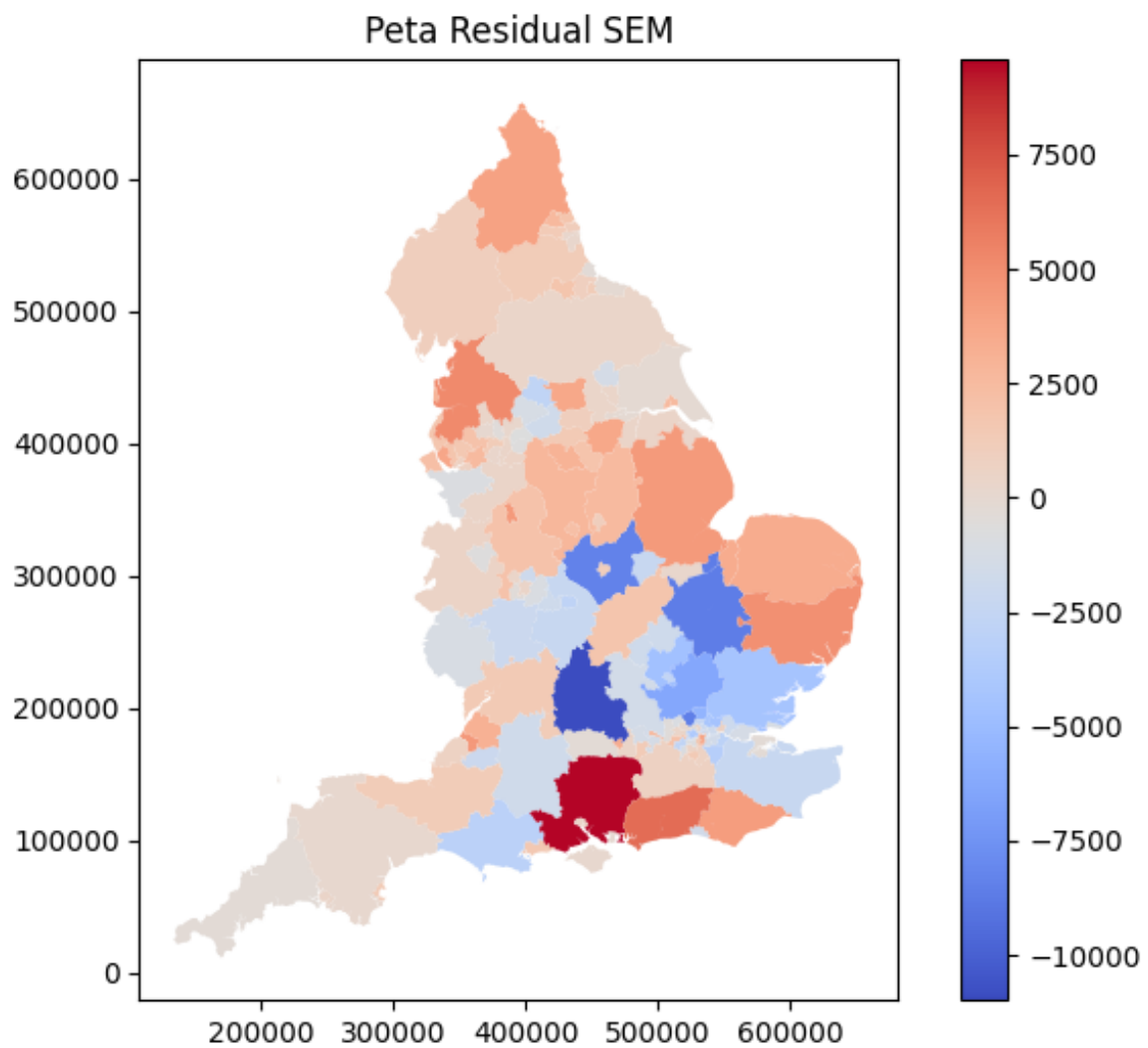
theory), di mana pendidikan meningkatkan kualitas dan kapasitas tenaga kerja dalam sektor pelayanan publik.

Selanjutnya, variabel *Financial* (Financial and Insurance) memiliki koefisien negatif sebesar -0.0249 dan tidak signifikan ($p\text{-value} = 0.733$). Nilai ini menunjukkan bahwa secara statistik, sektor keuangan dan asuransi tidak memiliki pengaruh yang bermakna terhadap variasi sektor kesehatan dan pekerjaan sosial. Ketidaksignifikanan ini dapat diinterpretasikan bahwa dinamika sektor keuangan mungkin tidak secara langsung mempengaruhi distribusi tenaga kerja kesehatan pada tingkat wilayah yang dianalisis, atau pengaruhnya telah terakomodasi melalui variabel lain dalam model.

Variabel *Info & Comm* (Information and Communication) memiliki koefisien negatif sebesar -0.6957 dan signifikan pada tingkat 1% ($p\text{-value} = 0.000$). Nilai $z\text{-statistic}$ sebesar -8.866 menunjukkan kekuatan pengaruh yang tinggi secara statistik. Koefisien negatif ini mengindikasikan bahwa peningkatan aktivitas sektor informasi dan komunikasi berkorelasi dengan penurunan proporsi sektor kesehatan dan pekerjaan sosial. Secara ekonomi regional, hal ini dapat mencerminkan adanya spesialisasi sektor, di mana wilayah yang berkembang pada sektor teknologi dan komunikasi cenderung memiliki struktur ekonomi yang lebih terfokus pada sektor tersier berbasis teknologi dibandingkan pelayanan kesehatan dan sosial.

Variabel *Real Estate* memiliki koefisien positif sebesar 3.0950 dan signifikan pada tingkat 1% ($p\text{-value} = 0.000$). Nilai $z\text{-statistic}$ sebesar 5.587 menunjukkan bahwa pengaruhnya signifikan secara statistik. Koefisien ini menunjukkan bahwa peningkatan sektor real estat berkorelasi dengan peningkatan sektor kesehatan dan pekerjaan sosial. Secara konseptual, perkembangan real estat sering kali mencerminkan pertumbuhan wilayah dan peningkatan kepadatan penduduk, yang pada gilirannya meningkatkan kebutuhan terhadap layanan kesehatan dan sosial.

Parameter spasial error (λ) memiliki nilai sebesar 0.2829 dan signifikan pada tingkat 1% ($p\text{-value} = 0.006$). Nilai ini menunjukkan adanya autokorelasi spasial yang signifikan pada komponen error. Dengan kata lain, terdapat faktor-faktor yang tidak teramati dalam model yang memiliki pola spasial dan mempengaruhi variabel dependen. Nilai λ yang positif mengindikasikan bahwa residual pada suatu wilayah cenderung berkorelasi positif dengan residual wilayah tetangganya. Keberadaan parameter λ yang signifikan membuktikan bahwa penggunaan model SEM lebih tepat dibandingkan model OLS klasik, karena model ini mampu mengakomodasi ketergantungan spasial yang sebelumnya terdeteksi melalui uji Moran's I pada residual.



Gambar 6: Peta Residual Model SEM

4.6 Perbandingan Model

Tabel 8: Perbandingan Kebaikan Model

Kriteria	OLS	SAR	SEM
AIC	2797.748	2799.583	2788.300
Log Likelihood	-1393.874	-1393.792	-1389.150

Model **SEM** memiliki nilai AIC terendah (2788.300) dibandingkan OLS (2797.748) dan SAR (2799.583), menegaskan bahwa SEM adalah model terbaik untuk data ini. Nilai koefisien λ (lambda) sebesar 0.283 yang signifikan menunjukkan adanya korelasi spasial yang moderat pada error yang berhasil ditangani oleh model.

4.7 Interpretasi

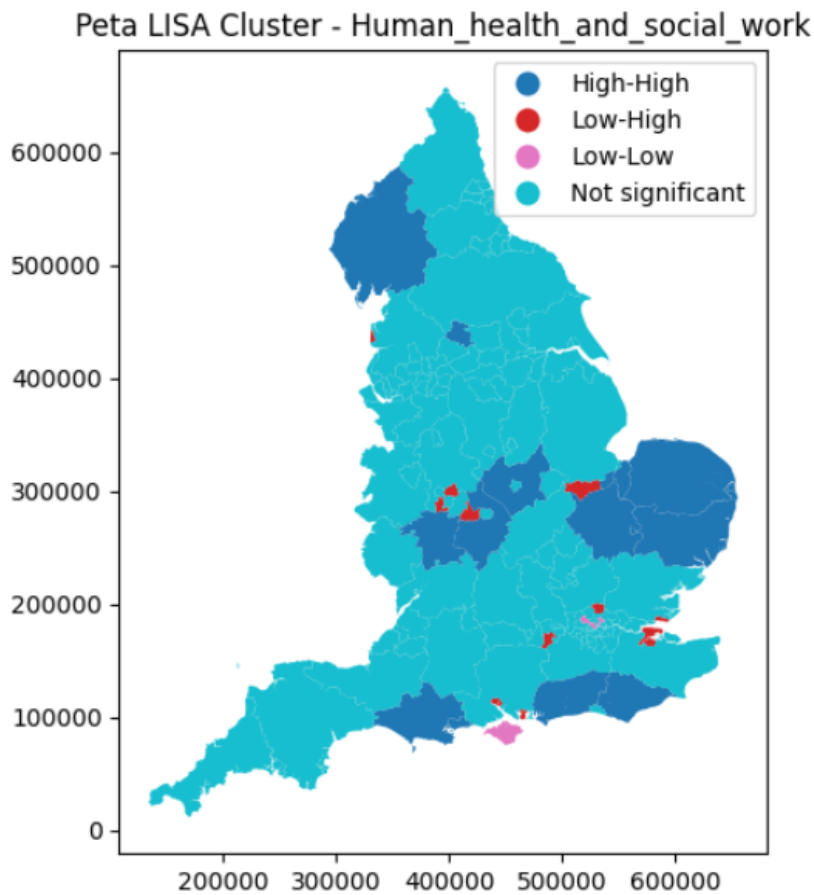
Dari model SEM, variabel `Education` dan `Real_estate` memiliki pengaruh positif signifikan terhadap `Human_health_and_social_work`. Artinya, wilayah dengan proporsi sektor pendidikan dan real estate yang tinggi cenderung memiliki proporsi pekerja sektor kesehatan yang tinggi pula. Sebaliknya, `Information_and_communication` memiliki pengaruh negatif signifikan.

BAB V

KESIMPULAN

Berdasarkan serangkaian analisis yang telah dilakukan menggunakan pendekatan regresi spasial untuk memodelkan proporsi tenaga kerja di sektor *Human Health and Social Work* di Inggris, diperoleh beberapa kesimpulan penting sebagai berikut:

1. **Eksistensi Dependensi Spasial:** Hasil uji Moran's I pada residual model OLS menunjukkan nilai statistik sebesar 4.145 dengan p-value < 0.001 . Hal ini mengindikasikan adanya autokorelasi spasial positif yang signifikan, yang berarti wilayah-wilayah dengan proporsi tenaga kerja kesehatan yang tinggi cenderung mengelompok berdekatan dengan wilayah sejenis, begitu pula sebaliknya. Temuan ini dikonfirmasi lebih lanjut melalui visualisasi peta *Local Indicators of Spatial Association* (LISA) Cluster Map (Gambar 7).



Gambar 7: Peta Klaster LISA (Local Indicators of Spatial Association)

Peta LISA di atas memperlihatkan adanya klaster *High-High* (wilayah dengan proporsi tenaga kerja kesehatan tinggi dikelilingi oleh wilayah serupa) yang signifikan di beberapa bagian Inggris, serta klaster *Low-Low* di area lainnya. Keberadaan pola pengelompokan ini menegaskan bahwa model OLS klasik yang mengasumsikan independensi antar observasi menjadi tidak valid dan bias.

2. **Pemilihan Model Terbaik:** Berdasarkan uji Lagrange Multiplier (LM), ditemukan bahwa parameter *LM-Error* signifikan ($p = 0.0002$), sedangkan *LM-Lag* tidak signifikan ($p = 0.6760$). Sesuai dengan prosedur seleksi model Anselin, model **Spatial Error Model (SEM)** dipilih sebagai model yang paling tepat. Keputusan ini didukung oleh perbandingan nilai *Akaike Information Criterion* (AIC), di mana model SEM memiliki nilai AIC terendah (**2788.30**) dibandingkan dengan model OLS (2797.75) dan model SAR (2799.58). Hal ini menunjukkan bahwa model SEM mampu memberikan keseimbangan terbaik antara kebaikan suai (*goodness of fit*) dan kompleksitas model.
3. **Determinan Tenaga Kerja Sektor Kesehatan:** Dari hasil estimasi model SEM, diidentifikasi faktor-faktor yang berpengaruh signifikan terhadap proporsi tenaga kerja di sektor kesehatan dan sosial:
 - **Sektor Pendidikan (*Education*):** Memiliki pengaruh positif yang sangat signifikan ($\beta = 1.028, p < 0.001$). Koefisien ini dapat diinterpretasikan bahwa setiap penambahan 1 orang tenaga kerja di sektor pendidikan diasosiasikan dengan peningkatan sekitar 1.03 orang tenaga kerja di sektor kesehatan, *ceteris paribus*. Hal ini menunjukkan adanya sinergi yang kuat antara sektor pendidikan dan kesehatan, di mana wilayah dengan basis pendidikan yang kuat cenderung mendukung ketersediaan tenaga kerja kesehatan yang lebih besar.
 - **Sektor Real Estat (*Real Estate*):** Berpengaruh positif signifikan ($\beta = 3.095, p < 0.001$). Artinya, setiap penambahan 1 pekerja di sektor real estat berhubungan dengan peningkatan sekitar 3.1 pekerja di sektor kesehatan. Hal ini mungkin mencerminkan hubungan antara pembangunan infrastruktur wilayah/pemukiman dengan penyediaan fasilitas layanan kesehatan.
 - **Sektor Informasi dan Komunikasi:** Memiliki pengaruh negatif signifikan ($\beta = -0.696, p < 0.001$). Ini mengindikasikan bahwa kenaikan 1 pekerja di sektor informasi dan komunikasi berasosiasi dengan penurunan sekitar 0.7 pekerja di sektor kesehatan. Pola ini menandakan adanya spesialisasi wilayah, di mana daerah yang menjadi pusat teknologi dan informasi cenderung memiliki proporsi tenaga kerja kesehatan yang lebih rendah secara relatif.
 - **Sektor Keuangan:** Tidak menunjukkan pengaruh yang signifikan secara statistik ($\beta = -0.025, p = 0.733$), sehingga perubahan pada jumlah pekerja sektor keuangan tidak memberikan dampak nyata terhadap jumlah pekerja sektor kesehatan.
4. **Signifikansi Parameter Error Spasial:** Parameter lambda (λ) pada model SEM bernilai 0.2829 dan signifikan pada taraf nyata 5% ($p = 0.006$). Signifikansi ini menunjukkan bahwa masih terdapat variabel-variabel lain yang tidak dimasukkan ke dalam model (*omitted variables*) yang memiliki pola spasial dan mempengaruhi distribusi tenaga kerja kesehatan. Model SEM berhasil menangkap dan mengontrol efek dari variabel-variabel laten tersebut melalui struktur error spasial, sehingga menghasilkan estimator koefisien regresi yang lebih efisien dibandingkan OLS.

Secara keseluruhan, penelitian ini menyimpulkan bahwa distribusi tenaga kerja di sektor kesehatan dan pekerjaan sosial di Inggris tidak terjadi secara acak, melainkan membentuk pola spasial yang dipengaruhi oleh struktur ekonomi wilayah (terutama pendidikan dan real estat) serta faktor-faktor spasial tak teramati lainnya. Implikasi kebijakan dari temuan ini adalah bahwa perencanaan ketenagakerjaan di sektor kesehatan harus mempertimbangkan karakteristik spasial dan interaksi antar wilayah, tidak hanya melihat kondisi internal masing-masing daerah secara terisolasi.

DAFTAR PUSTAKA

- [1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.
- [2] Rey, S. J., & Anselin, L. (2010). PySAL: A Python Library of Spatial Analytical Methods. *Review of Regional Studies*, 37, 5-27.
- [3] Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1), 234-240.
- [4] Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- [5] Office for National Statistics (ONS). (2020). Labour Market Profile - Nomis - Official Labour Market Statistics.

LAMPIRAN

Seluruh data penelitian, kode program (*source code*), serta hasil keluaran pendukung lainnya dapat diakses melalui tautan Google Drive berikut:

[https://drive.google.com/drive/folders/
1ut6zhfMA0k8HP2EK1fGR5h6ZpF6eLB60?usp=sharing](https://drive.google.com/drive/folders/1ut6zhfMA0k8HP2EK1fGR5h6ZpF6eLB60?usp=sharing)