



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 Matriks Pembobot Spasial (Spatial Weights Matrix)	6
2.1.1 Rook Contiguity	6
2.1.2 Queen Contiguity	6
2.1.3 Bishop Contiguity	7
2.2 Indeks Moran (Moran's I)	7
2.3 Uji Lagrange Multiplier (LM)	7
2.4 Spatial Lag Model (SAR)	8
2.5 Spatial Error Model (SEM)	8
BAB III METODE PENELITIAN	9
3.1 Data dan Variabel	9
3.2 Tahapan Analisis	9
BAB IV ANALISIS DAN PEMBAHASAN	10
4.1 Analisis Deskriptif dan Eksplorasi Data (EDA)	10
4.1.1 Statistik Deskriptif	10
4.1.2 Visualisasi Data	10
4.1.3 Peta Sebaran	12
4.2 Hasil Regresi OLS (Global)	13
4.3 Diagnostik Dependensi Spasial	13
4.4 Hasil Regresi Spasial (SEM)	15
4.5 Perbandingan Model	15
4.6 Interpretasi	16
BAB V KESIMPULAN	17
DAFTAR PUSTAKA	18
LAMPIRAN	19

Daftar Tabel

1	Statistik Deskriptif Variabel Penelitian	10
2	Ringkasan Hasil Regresi OLS	13
3	Hasil Uji Diagnostik Spasial	14
4	Ringkasan Hasil Spatial Error Model (SEM)	15
5	Perbandingan Kebaikan Model	15

Daftar Gambar

1	Histogram Distribusi Variabel	11
2	Boxplot Variabel (Deteksi Outlier)	11
3	Matriks Korelasi Antar Variabel	12
4	Peta Sebaran Tenaga Kerja Sektor Kesehatan (Quantile)	13
5	Peta Residual: OLS vs SAR vs SEM	14
6	Visualisasi Autokorelasi Spasial Lokal (LISA)	15

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 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:

$$W = [w_{ij}] \quad (1)$$

dengan:

$$w_{ij} = \begin{cases} 1, & \text{jika wilayah } i \text{ bertetangga dengan wilayah } j \\ 0, & \text{lainnya} \end{cases} \quad (2)$$

Untuk menjaga konsistensi skala, dilakukan normalisasi baris:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}} \quad (3)$$

Sehingga setiap baris memiliki jumlah bobot sebesar 1.

2.1.1 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:

$$w_{ij}^{rook} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi sisi} \\ 0, & \text{lainnya} \end{cases} \quad (4)$$

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

2.1.2 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.

$$w_{ij}^{queen} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi sisi atau titik sudut} \\ 0, & \text{lainnya} \end{cases} \quad (5)$$

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

2.1.3 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.

$$w_{ij}^{bishop} = \begin{cases} 1, & \text{jika } i \text{ dan } j \text{ berbagi titik sudut saja} \\ 0, & \text{lainnya} \end{cases} \quad (6)$$

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

2.2 Indeks Moran (Moran's I)

Indeks Moran digunakan untuk mengukur autokorelasi spasial global.

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (7)$$

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

2.3 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.

2.4 Spatial Lag Model (SAR)

Model SAR mengasumsikan bahwa nilai variabel dependen di suatu lokasi dipengaruhi oleh nilai variabel dependen di lokasi tetangganya.

$$y = \rho W y + X\beta + \epsilon$$

di mana ρ adalah parameter lag spasial.

2.5 Spatial Error Model (SEM)

Model SEM mengasumsikan bahwa ketergantungan spasial terdapat pada residual, yang mungkin disebabkan oleh variabel tak teramati yang berkorelasi spasial.

$$y = X\beta + u$$

$$u = \lambda W u + \epsilon$$

di mana λ adalah parameter error spasial.

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`).

Seluruh variabel dilakukan standardisasi (Z-score) sebelum analisis untuk memudahkan interpretasi dan komputasi.

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

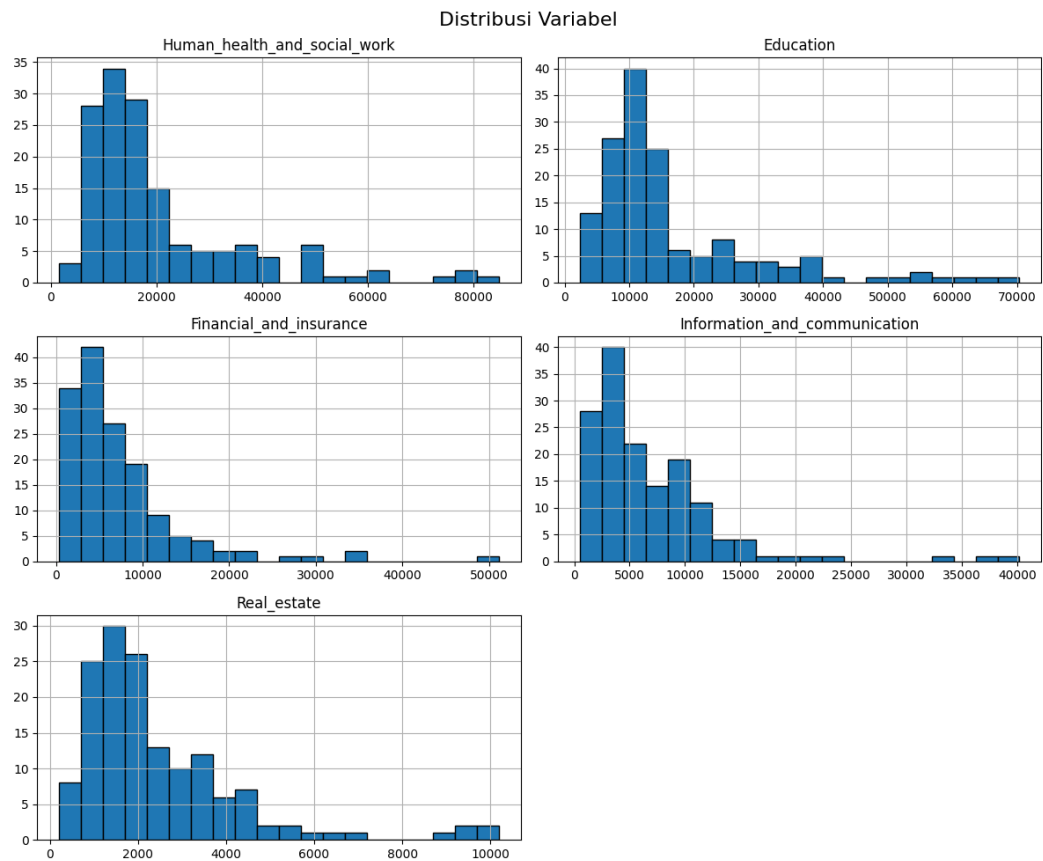
Tabel 1 menyajikan ringkasan statistik untuk variabel-variabel yang digunakan.

Tabel 1: Statistik Deskriptif Variabel Penelitian

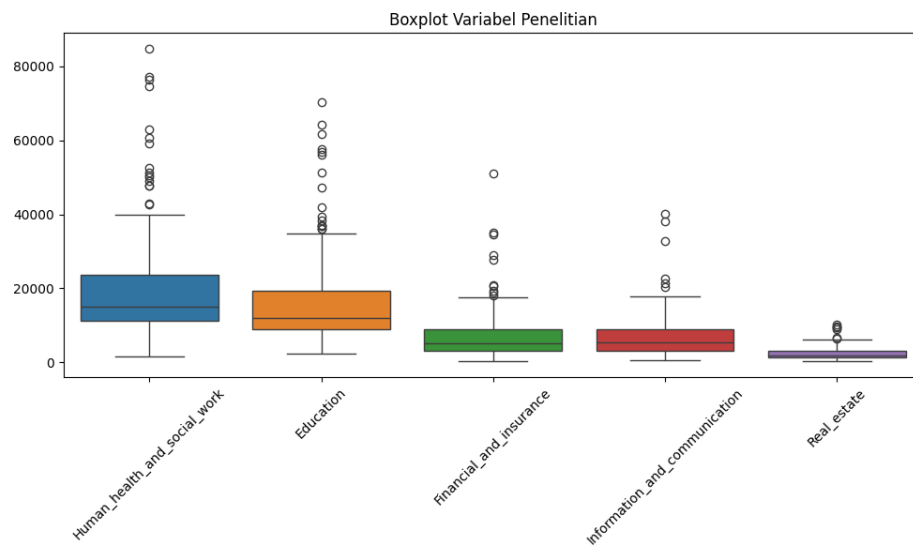
Variable	Count	Mean	Std	Min	25%	50%	75%	Max	Skew	Kurtosis
Health & Social	149	20947.9	15772.2	1585	11219	15118	23558	84796	1.92	3.72
Education	149	16712.7	13157.9	2386	8827	11951	19418	70363	2.00	3.99
Financial	149	7408.4	7010.5	387	3196	5190	9028	51117	2.97	12.50
Info & Comm	149	6874.8	6152.1	518	3018	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.

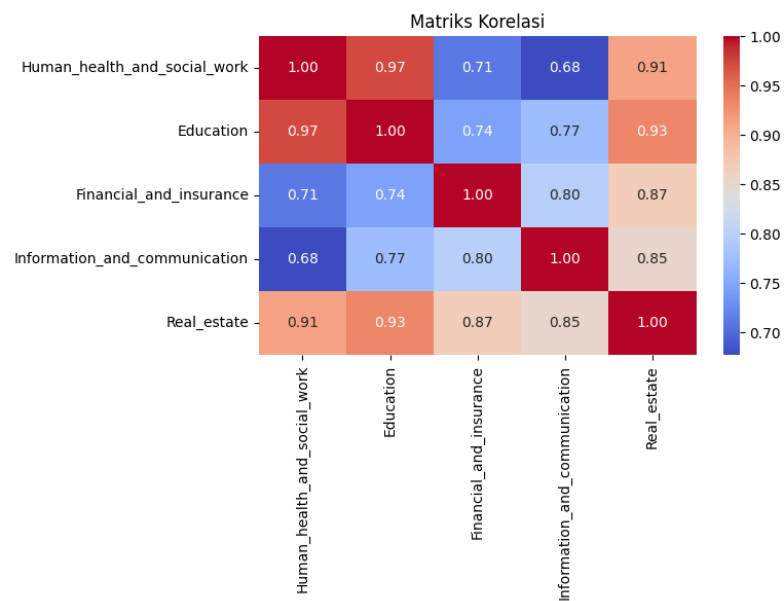


Gambar 1: Histogram Distribusi Variabel



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.

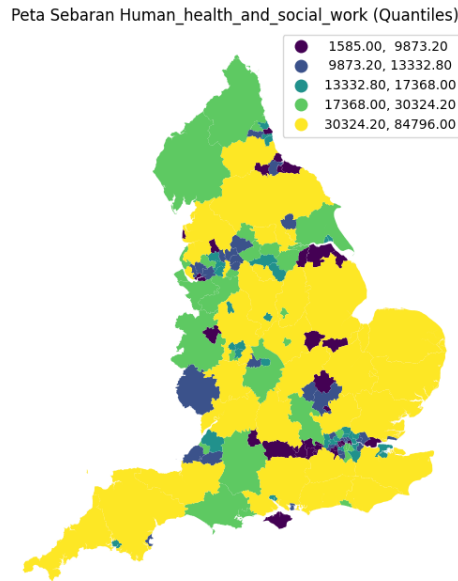


Gambar 3: Matriks Korelasi Antar Variabel

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). Ringkasan hasilnya adalah sebagai berikut:

Tabel 2: Ringkasan Hasil Regresi OLS

Variabel	Koefisien	Std. Error	t-Statistic	Prob.
CONSTANT	-0.0000	0.0148	-0.0000	1.000
Education	0.8892	0.0447	19.912	0.000
Financial	-0.0234	0.0328	-0.714	0.476
Info & Comm	-0.2829	0.0292	-9.705	0.000
Real Estate	0.3422	0.0639	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 Diagnostik Dependensi Spasial

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

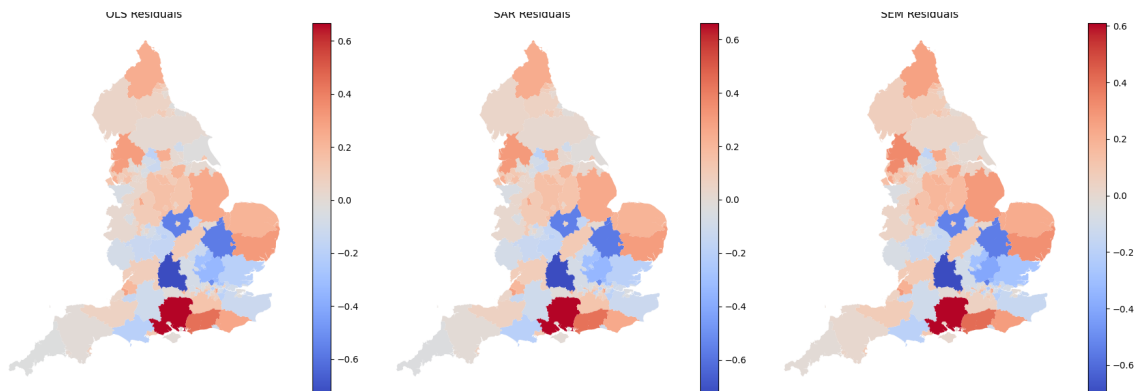
Tabel 3: Hasil Uji Diagnostik Spasial

Uji	Nilai Statistik	Probabilitas (p-value)
Moran's I (error)	4.145	0.0000
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

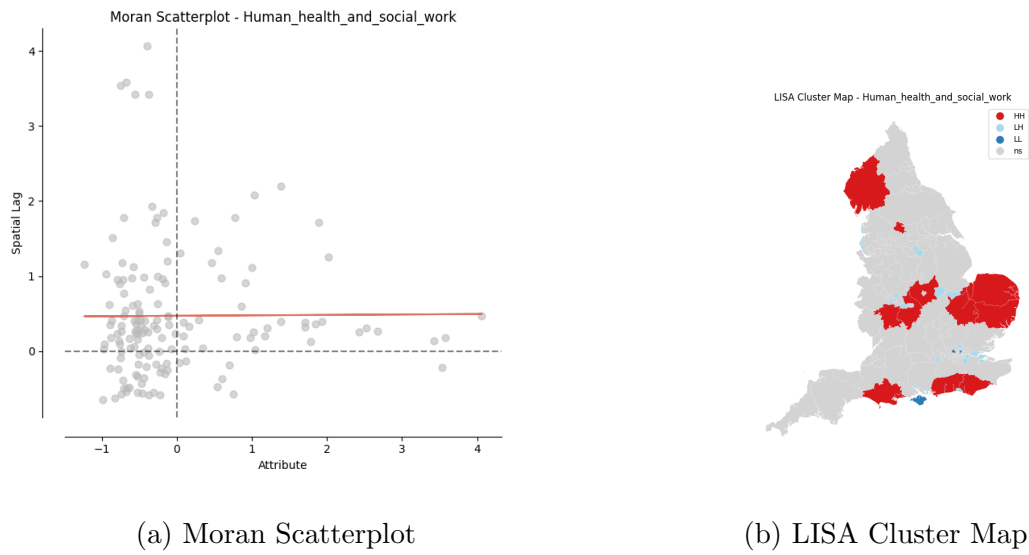
Berdasarkan Tabel 3:

- 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



Gambar 6: Visualisasi Autokorelasi Spasial Lokal (LISA)

Gambar 6 memperlihatkan adanya pengelompokan nilai tinggi (High-High) dan rendah (Low-Low) yang signifikan secara statistik di beberapa wilayah, yang mengonfirmasi adanya dependensi spasial lokal.

4.4 Hasil Regresi Spasial (SEM)

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

Tabel 4: Ringkasan Hasil Spatial Error Model (SEM)

Variabel	Koefisien	Std. Error	z-Statistic	Prob.
CONSTANT	0.0001	0.0197	0.005	0.996
Education	0.8580	0.0464	18.486	0.000
Financial	-0.0111	0.0324	-0.342	0.733
Info & Comm	-0.2714	0.0306	-8.866	0.000
Real Estate	0.3616	0.0647	5.587	0.000
Lambda (λ)	0.2829	0.1028	2.752	0.006

4.5 Perbandingan Model

Tabel 5: Perbandingan Kebaikan Model

Kriteria	OLS	SAR	SEM
AIC	-81.72	-79.88	-91.17
Log Likelihood	45.86	45.94	50.58

Model **SEM** memiliki nilai AIC terendah (-91.17) dibandingkan OLS (-81.72) dan SAR (-79.88), 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.

4.6 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 analisis yang dilakukan, dapat disimpulkan bahwa:

1. Terdapat autokorelasi spasial yang signifikan pada residual model OLS, sehingga model OLS klasik tidak sepenuhnya memadai.
2. Berdasarkan uji Lagrange Multiplier, model yang paling sesuai adalah **Spatial Error Model (SEM)**.
3. Model SEM memberikan performa terbaik dengan AIC terendah (-91.17).
4. Faktor Pendidikan dan Real Estate berpengaruh positif signifikan, sedangkan Informasi & Komunikasi berpengaruh negatif signifikan terhadap variabel respon.

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 Python (*run_spatial_analysis.py*), serta hasil keluaran pendukung lainnya telah diunggah ke repositori GitHub:

<https://github.com/ammarnfi/TK2>