

Pemetaan Spasial Kesenjangan Akses Pendidikan di Indonesia Tahun 2023

Spatial Mapping of Education Access Disparities in Indonesia in 2023

Kelompok 5 - Official Statistics (B)

Ferdinand Sitompul (2206051576), Kartika Rizkia Zuhrah (2206027993),

Kayla Zahira Amadya (2206053890), Najwa Putri Faradila (2206051355),

Ester Oktafia (2206810830), Naufal Elban Musyaffa (2206053865),

Muhammad Binar Raffi (2306228005)

Abstract.

This study analyzes the gap in access to education between Java and other islands in Indonesia, as part of strategies to achieve Sustainable Development Goal 4, which is to provide quality education for all. Using comparative analysis and Geographically Weighted Regression (GWR), the study identifies significant regional variations in factors influencing access to education. High school graduation rates and poverty emerge as key determinants, but their impacts differ across regions. These findings emphasize the importance of focused education programs corresponded with specific regional conditions.

Keywords: Quality Education, Education Access, Disparities, Spatial Mapping, Geographically Weighted Regression

Abstrak.

Studi ini menganalisis kesenjangan akses pendidikan antara pulau Jawa dan luar Jawa di Indonesia, sebagai upaya untuk mencapai Tujuan Pembangunan Berkelanjutan 4, yaitu menyediakan pendidikan berkualitas untuk semua. Dengan menggunakan analisis komparatif dan *Geographically Weighted Regression* (GWR), studi ini mengidentifikasi variasi regional yang signifikan dalam faktor-faktor yang memengaruhi akses pendidikan. Tingkat penyelesaian pendidikan jenjang Sekolah Menengah Atas dan persentase kemiskinan penduduk menjadi faktor utama, tetapi dampaknya berbeda di setiap wilayah. Hasil ini menekankan pentingnya program pendidikan yang terfokus sesuai dengan kondisi regional tertentu.

Kata Kunci: Pendidikan Berkualitas, Akses Pendidikan, Kesenjangan, Pemetaan Spasial, *Geographically Weighted Regression*

Pendahuluan

Pendidikan merupakan salah satu pilar utama dalam pembangunan berkelanjutan, sebagaimana tercantum dalam Tujuan Pembangunan Berkelanjutan (TPB) nomor 4, yaitu memastikan pendidikan yang inklusif, berkualitas, serta mendorong kesempatan belajar sepanjang

hayat untuk semua. Di Indonesia, tantangan dalam mencapai tujuan ini masih cukup kompleks, khususnya terkait dengan kesenjangan akses pendidikan antara wilayah Jawa dan luar Jawa.

Wilayah Jawa yang memiliki infrastruktur serta sumber daya lebih baik cenderung menunjukkan tingkat partisipasi sekolah yang lebih tinggi dibandingkan wilayah luar Jawa yang relatif tertinggal. Perbedaan ini dapat disebabkan oleh berbagai faktor sosial-ekonomi, seperti tingkat pengangguran, pendapatan per kapita, persentase penduduk miskin, serta ketimpangan gender yang memengaruhi kemampuan masyarakat dalam mengakses pendidikan (Khoiriyani, Dani, & Ansyah, 2023).

Studi sebelumnya menunjukkan bahwa kesenjangan akses pendidikan antara wilayah Jawa dan luar Jawa berimplikasi signifikan terhadap kualitas layanan pendidikan. Penelitian yang dilakukan oleh Fauziah Khoiriyani, Fatimatuz Zahro Diah Putri Dani, dan Rahmat Husein Andri Ansyah (2023) menemukan adanya kesenjangan kualitas layanan Madrasah Aliyah Negeri (MAN) di kedua wilayah tersebut. Kesenjangan ini menjadi penghambat dalam upaya mewujudkan akses pendidikan yang merata dan berkeadilan.

Selain itu, Abdulloh dan Yasin (2024) menyimpulkan bahwa kesenjangan akses pendidikan berkualitas di wilayah Sangatta Selatan, Kalimantan Timur berdampak signifikan terhadap mobilitas sosial. Anak-anak dari keluarga kaya memiliki peluang pendidikan yang lebih baik dibandingkan dengan mereka yang berasal dari keluarga miskin, sehingga memperkuat siklus kemiskinan dan ketimpangan sosial-ekonomi. Faktor ekonomi keluarga berperan penting dalam mendukung efektivitas pendidikan sebagai sarana mobilitas sosial, di mana keluarga mampu memberikan dukungan lebih besar, sementara keluarga kurang mampu menghadapi berbagai hambatan.

Tanpa intervensi yang tepat, ketimpangan ini dapat memperdalam kesenjangan sosial dan ekonomi yang berpotensi menghambat pencapaian Tujuan Pembangunan Keberlanjutan di Indonesia secara keseluruhan. Oleh karena itu, penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai kondisi akses pendidikan di Indonesia, khususnya dalam konteks kesenjangan antara wilayah Jawa dan luar Jawa dengan metode uji t dua sampel independen. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi akses pendidikan di berbagai provinsi di Indonesia dengan metode *Geographically Weighted Regression* (GWR). Hasil penelitian ini diharapkan dapat menjadi dasar dalam merumuskan kebijakan yang lebih efektif dan inklusif untuk mengurangi kesenjangan pendidikan serta mendukung tercapainya pembangunan berkelanjutan.

Metode

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) pada tahun 2023 dengan wilayah penelitian 34 provinsi di Indonesia. Data yang digunakan terdiri dari berbagai indikator sosial dan ekonomi yang memengaruhi Angka Partisipasi Sekolah (APS), yang dijadikan sebagai variabel dependen (Y) dalam penelitian ini.

Angka Partisipasi Sekolah (APS) sebagai variabel dependen akan dianalisis dengan mempertimbangkan pengaruh dari berbagai variabel independen yang berkaitan dengan kondisi ekonomi, sosial, dan infrastruktur pendidikan. Beberapa variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 sebagai berikut.

Tabel 1. Variabel yang Digunakan	
Nama Variabel	Keterangan
Angka Partisipasi Sekolah pada Jenjang Pendidikan SMA (APS SMA)	Persentase jumlah penduduk usia sekolah yang terdaftar di sekolah dibandingkan dengan jumlah penduduk yang memiliki usia sekolah yang sesuai.
Persentase Penduduk Miskin (permiskin)	Persentase jumlah penduduk yang hidup di bawah garis kemiskinan, yang dipengaruhi oleh faktor ekonomi dan sosial dalam wilayah penelitian.
Indeks Ketimpangan Gender (IKG)	Indeks yang mengukur kesenjangan antara laki-laki dan perempuan dalam berbagai aspek kehidupan, termasuk pendidikan dan tenaga kerja.
Tingkat Pengangguran Terbuka (TPT)	Persentase penduduk usia kerja yang tidak memiliki pekerjaan meskipun sedang mencari kerja.
Produk Domestik Regional Bruto per Kapita atas Dasar Harga Berlaku (PDRB)	Ukuran ekonomi regional yang dihitung per kapita, mencerminkan tingkat kemakmuran ekonomi di setiap provinsi.
Indeks Pembangunan Manusia (IPM)	Indeks yang mengukur kualitas kehidupan di suatu daerah, meliputi kesehatan, pendidikan, dan standar hidup yang lebih tinggi.
Jumlah SMA	Jumlah sekolah menengah atas (SMA) yang ada di suatu provinsi.
Persentase Jumlah Guru (perguru)	Persentase jumlah guru di suatu provinsi terhadap jumlah guru di seluruh Indonesia.
Tingkat Penyelesaian Pendidikan menurut Jenjang Pendidikan SMA (TPP)	Persentase jumlah siswa yang berhasil menyelesaikan pendidikan pada jenjang SMA di tiap provinsi.

Untuk mencapai tujuan pemahaman yang mendalam mengenai kondisi akses pendidikan di Indonesia, khususnya wilayah Jawa dan luar Jawa, digunakan beberapa metode dalam penelitian ini, yaitu uji *t* dua sampel independen, regresi linier berganda, dan *Geographically Weighted Regression* (GWR). Uji *t* dua sampel independen digunakan untuk menguji apakah terdapat perbedaan signifikan mengenai rata-rata rasio anak usia aktif SMA (16-18 tahun) dengan jumlah SMA tahun 2023. Kemudian, metode analisis regresi linier berganda serta *Geographically Weighted Regression* (GWR) digunakan sebagai analisis lanjutan untuk memetakan hubungan faktor sosial-

ekonomi terhadap kondisi pendidikan di berbagai provinsi di Indonesia dengan mempertimbangkan aspek spasial tiap provinsi tersebut.

1. Uji t Dua Sampel Independen

Uji t dua sampel independen, atau *Independent-Samples t-Test*, digunakan untuk menguji apakah terdapat perbedaan yang signifikan secara statistik antara rata-rata dua kelompok sampel yang independen. Pengujian ini sering digunakan untuk membandingkan kelompok dengan karakteristik yang berbeda, seperti membandingkan skor ujian siswa dari dua sekolah yang berbeda. Secara matematis, hipotesis nol (H_0) dalam uji t menyatakan bahwa tidak ada perbedaan antara rata-rata kedua kelompok ($\mu_1 = \mu_2$), sementara hipotesis alternatif (H_1) menyatakan adanya perbedaan ($\mu_1 \neq \mu_2$). Nilai statistik uji t dihitung dengan persamaan (1) berikut.

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (1)$$

Keterangan:

\bar{X} dan \bar{Y} : rata-rata dari kelompok 1 dan 2

S_1^2 dan S_2^2 : variansi dari kelompok 1 dan 2

n_1 dan n_2 : ukuran sampel dari kelompok 1 dan 2

2. Regresi Linier Berganda

Regresi linier berganda digunakan untuk menganalisis hubungan antara satu variabel dependen Y dan beberapa variabel independen X . Model matematis dari regresi linier berganda adalah sebagai berikut.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (2)$$

Keterangan:

Y : variabel dependen.

β_0 : intersep (konstanta).

β_1, \dots, β_n : koefisien regresi yang menunjukkan pengaruh masing-masing X_i terhadap Y .

X_1, \dots, X_n : variabel independen.

ϵ : *error* (gangguan) yang diasumsikan terdistribusi normal dengan rata-rata 0 dan varians konstan.

Setelah membentuk pemodelan seperti pada persamaan (2), perlu dilakukan pengecekan asumsi model regresi linier berganda sebagai berikut.

a. Homoskedastisitas Spasial

Variansi residual harus konstan atau tidak bergantung pada nilai prediksi. Uji homoskedastisitas spasial bertujuan untuk memeriksa apakah hubungan antara variabel dependen dan independen berbeda antarwilayah. Adanya heterogenitas spasial menunjukkan bahwa model global atau model regresi linier berganda mungkin tidak cukup akurat karena adanya variasi lokal. Uji ini dapat dilakukan menggunakan uji Breusch-Pagan dengan persamaan matematis sebagai berikut.

$$BP = \frac{1}{2} \vec{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \vec{f} \quad (3)$$

Dengan elemen vektor \vec{f} adalah $f_i = \left(\frac{e_i^2}{\sigma^2} - 1 \right)$, di mana e_i^2 adalah *least square residual* untuk observasi ke- i , σ^2 adalah variansi residual, dan matriks \mathbf{Z} adalah matriks yang memuat vektor berdistribusi normal standar baku untuk setiap observasi dengan ukuran $n \times (p + 1)$.

b. Normalitas Residual

Residual harus berdistribusi normal. Untuk menguji normalitas, dapat digunakan uji Shapiro-Wilk atau Q-Q plot. Secara matematis, normalitas dapat diuji dengan menghitung nilai p dari distribusi normal residual.

$$W = \frac{(\sum_{i=1}^n (e_i - \bar{e})^2)^2}{\sum_{i=1}^n (e_i - \hat{e})^2} \quad (4)$$

Jika W pada persamaan (4) mendekati 1, maka residual cenderung berdistribusi normal.

c. Independensi Residual (Autokorelasi)

Residual harus bebas dari autokorelasi, terutama saat data merupakan data runtun waktu (*time series*). Untuk menguji autokorelasi, dapat digunakan uji Durbin-Watson seperti pada persamaan (5).

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (5)$$

dengan e_t adalah residual pada waktu t . Nilai Durbin-Watson berkisar antara 0 hingga 4, dengan ketentuan sebagai berikut.

- Nilai sekitar 2 menunjukkan tidak ada autokorelasi.
- Nilai mendekati 0 menunjukkan autokorelasi positif.
- Nilai mendekati 4 menunjukkan autokorelasi negatif.

d. Multikolinieritas

Variabel independen pada model tidak boleh memiliki korelasi yang tinggi satu sama lain. Untuk mendeteksi multikolinieritas, digunakan *Variance Inflation Factor* (VIF). Rumus VIF untuk setiap variabel independen X_i adalah sebagai berikut.

$$VIF_i = \frac{1}{1 - R_i^2} \quad (6)$$

dengan R_i^2 adalah koefisien determinasi dari regresi X_i terhadap semua variabel independen lainnya. VIF yang lebih besar dari 10 menunjukkan multikolinieritas yang tinggi.

3. Geographically Weighted Regression (GWR)

Geographically Weighted Regression atau GWR merupakan pemodelan yang mempertimbangkan pengaruh lokasi observasi sehingga setiap lokasi memiliki nilai parameter model yang berbeda-beda sebanyak $n \times (p + 1)$ yang diestimasi dengan n sebagai jumlah lokasi pengamatan dan p adalah jumlah variabel. Pengaruh lokasi yang dimaksud adalah adanya heterogenitas spasial yang diuji setelah melakukan regresi linier berganda.

Model matematis *Geographically Weighted Regression* (GWR) adalah sebagai berikut.

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (7)$$

Keterangan:

Y_i : variabel dependen pada lokasi observasi ke- i

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

(u_i, v_i) : koordinat lokasi observasi ke- i

$\beta_0(u_i, v_i)$: konstanta model GWR pada lokasi observasi ke- i

$\beta_k(u_i, v_i)$: parameter model GWR ke- k pada lokasi observasi ke- i

ε_i : galat pada lokasi observasi ke- i

Suatu model GWR melibatkan matriks pembobot spasial, yaitu matriks diagonal berukuran $(n \times n)$ dengan elemen diagonal w_{ij} sebagai bobot untuk titik observasi ke- j terhadap titik observasi ke- i (u_i, v_i) .

$$W(u_i, v_i) = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{(n-1)1} & w_{(n-1)2} & \dots & w_{(n-1)(n-1)} & w_{(n-1)n} \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix} \quad (8)$$

Nilai pembobot spasial w_{ij} didapatkan menggunakan fungsi pembobot spasial. Semakin jauh titik observasi dengan titik dibangunnya model, semakin kecil nilai bobot yang diperoleh. Misalkan d_{ij} menyatakan jarak *Euclidean* titik observasi i dan j dengan formula berikut.

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (9)$$

Diperoleh bobot titik observasi ke- j untuk model pada lokasi ke- i adalah $w_j(u_i, v_i) = K(d_{ij})$ dengan $K(d_{ij})$ adalah fungsi pembobot spasial.

Pendugaan parameter model GWR dilakukan dengan metode *Weighted Least Square* (WLS), yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi observasi. Pendugaan parameter dilakukan dengan meminimumkan Jumlah Kuadrat Galat Terboboti (JKGT) dengan bentuk matematis sebagai berikut.

$$JKGT = \sum_{i=1}^n w_i(u_i, v_i) \varepsilon_i^2 = \boldsymbol{\varepsilon}' \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} \quad (10)$$

dengan $\boldsymbol{\varepsilon}' \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} = \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - 2\boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i)$.

Taksiran $\hat{\boldsymbol{\beta}}(u_i, v_i)$ dapat diperoleh ketika JKGT mencapai nilai minimum, yaitu saat turunan JKGT terhadap $\boldsymbol{\beta}(u_i, v_i)$ sama dengan nol.

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}' \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}(u_i, v_i)} &= 0 \\ \frac{\partial \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - 2\boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i)}{\partial \boldsymbol{\beta}(u_i, v_i)} &= 0 \\ 0 - 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= 0 \\ 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= 0 \\ 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= 2\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ \boldsymbol{\beta}(u_i, v_i) &= [\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \end{aligned} \quad (11)$$

Diperoleh penduga parameter GWR pada lokasi pengamatan ke- i (11) dengan $\mathbf{W}(u_i, v_i)$ adalah matriks pembobot spasial.

Fungsi pembobot spasial digunakan untuk menentukan elemen matriks pembobot spasial, di mana matriks tersebut akan digunakan untuk melakukan pendugaan parameter dalam model GWR. Terdapat dua fungsi pembobot spasial yang umum digunakan, yaitu fungsi *Kernel Gaussian* dan *Kernel Bisquare*.

a. Fungsi *Kernel Gaussian* memiliki formula sebagai berikut.

$$K_G(d_{ij}) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right] \quad (12)$$

b. Fungsi *Kernel Bisquare* memiliki formula sebagai berikut.

$$K_B(d_{ij}) = \left(1 - \frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right) \quad (13)$$

Pada fungsi *Kernel Bisquare*, bobot titik observasi yang jaraknya melebihi nilai *bandwidth* optimum adalah nol.

Keterangan:

d_{ij} : jarak *Euclidean* antara lokasi dibangunnya model GWR (i) dan lokasi observasi lainnya (j)

b : nilai *bandwidth* optimum

Bandwidth atau lebar jendela optimum adalah suatu jarak optimum yang memungkinkan respon suatu wilayah berpengaruh kuat terhadap wilayah lainnya. Penentuan nilai *bandwidth* optimum dapat menggunakan kalibrasi nilai *cross-validation* yang diformulasikan sebagai berikut.

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (14)$$

dengan y_i adalah variabel dependen ke- i dan $\hat{y}_{\neq i}$ adalah nilai estimasi untuk variabel dependen selain ke- i . Kalibrasi dilakukan dengan cara memasukkan berbagai nilai *bandwidth* ke dalam rumus *CV* sampai ditemukan nilai yang terkecil. Nilai *bandwidth* (b) dikatakan optimum jika nilai *CV* minimum.

Fungsi pembobot spasial yang baik akan memenuhi tiga sifat berikut.

- $K(0) = 1$;
- $\lim_{d_{ij} \rightarrow \infty} K(d_{ij}) = 0$; dan
- $K(d_{ij})$ adalah fungsi yang monoton turun.

4. Uji Signifikansi Parameter Model GWR

Untuk mengetahui variabel yang berpengaruh secara signifikan di suatu lokasi, akan dilakukan uji hipotesis dengan hipotesis nol $H_0: \beta_k(u_i, v_i) = 0$ dan hipotesis alternatifnya adalah $H_1: \beta_k(u_i, v_i) \neq 0$. Statistik uji yang digunakan merupakan bentuk perluasan dari uji t sebagai berikut.

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{se(\hat{\beta}_k(u_i, v_i))} \quad (15)$$

dengan $se(\hat{\beta}_k(u_i, v_i)) = \sqrt{C C^T \sigma^2}$ dan $C = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i)$.

Hipotesis nol ditolak jika $|t_k(u_i, v_i)| > t_{n-m-1, \frac{\alpha}{2}}$ atau $p\text{-value} < \alpha$.

Keterangan:

$se(\hat{\beta}_k(u_i, v_i))$: *standard error* untuk pendugaan parameter variabel dependen ke- k pada lokasi ke- i .

n : jumlah titik lokasi yang dianalisis.

m : jumlah variabel bebas.

5. Penilaian Kualitas Model GWR

Untuk membandingkan kualitas suatu model GWR, dapat diukur menggunakan skor *Akaike's Information Criterion* (AIC) dan koefisien determinasi.

a. Akaike's Information Criterion (AIC)

AIC merupakan sebuah ukuran relatif kualitas suatu model statistik terhadap model statistik lain. Model dengan skor AIC lebih kecil dianggap sebagai model yang lebih baik.

Skor AIC suatu model dapat dihitung dengan formula berikut.

$$AIC = \exp\left(\frac{2k}{n}\right) \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n} \quad (16)$$

Keterangan:

k : jumlah parameter.

n : jumlah observasi.

$\hat{\epsilon}_i$: galat pada observasi ke- i .

b. Koefisien Determinasi (R^2)

Koefisien determinasi merupakan suatu ukuran variasi variabel dependen yang dapat diprediksi oleh variabel independen. Semakin besar skor R^2 suatu model, maka model tersebut dianggap semakin baik. Koefisien determinasi dapat dihitung dengan formula berikut.

$$R^2 = 1 - \frac{JKG}{JKT} = 1 - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

Keterangan:

JKG : jumlah kuadrat galat

JKT : jumlah kuadrat total

y_i : variabel dependen pada observasi ke- i

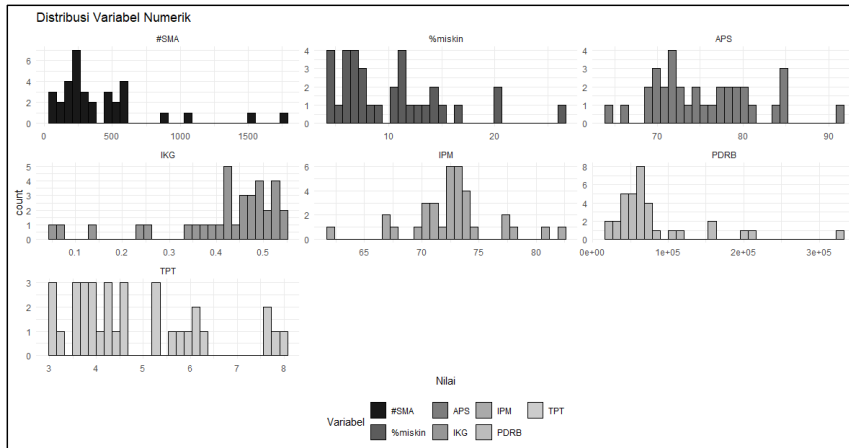
\bar{y} : mean variabel dependen

Hasil dan Analisis

Berdasarkan hasil statistik deskriptif pada Tabel 2, terdapat 34 provinsi yang diamati. Angka Partisipasi Sekolah (APS) memiliki nilai minimum 64,15% dan maksimum 91,17%, menunjukkan variasi yang cukup tinggi antar provinsi. Persentase penduduk miskin berkisar antara 4,25% hingga 26,03%, dengan rata-rata 10,08%, mengindikasikan adanya kesenjangan tingkat kemiskinan. Indeks Ketimpangan Gender (IKG) cukup kecil dengan nilai rata-rata 0,41, menunjukkan ketimpangan yang relatif rendah. Tingkat Pengangguran Terbuka (TPT) memiliki median 4,41%, dengan nilai maksimum 9,52%. Produk Domestik Regional Bruto (PDRB) sangat bervariasi, mulai dari 16.870 hingga 322.615, mencerminkan perbedaan ekonomi yang signifikan. Indeks Pembangunan Manusia (IPM) bervariasi dari 62,25 hingga 82,46, dengan median 72,78. Jumlah SMA tiap provinsi juga sangat beragam, mulai dari 68 hingga 1.768, menunjukkan perbedaan kapasitas pendidikan antar wilayah.

Tabel 2. Statistika Deskriptif Data

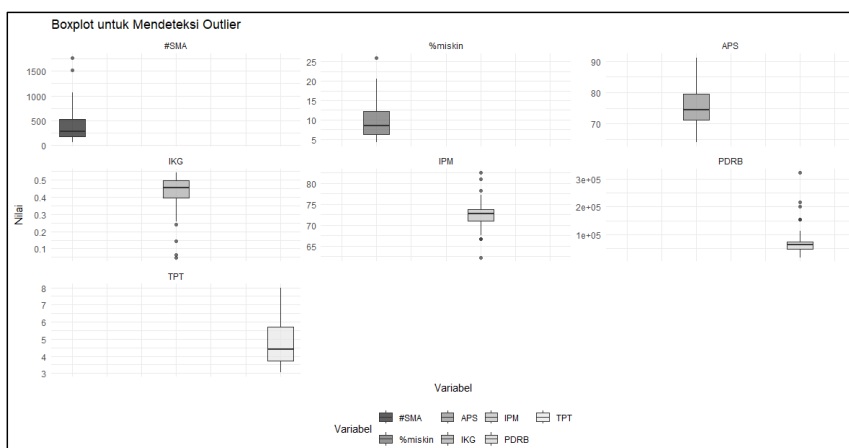
Statistika Deskriptif	APS SMA (%)	Persentase Kemiskinan (%)	IKG	TPT (%)	PDRB	IPM	Jumlah SMA
Min.	64.15	4.250	0.0490	3.040	16870	62.25	68
Median	74.58	8.425	0.4550	4.415	63934	72.78	278
Mean	75.50	10.089	0.4149	4.806	80149	72.62	424.90
Max.	91.17	26.030	0.5410	9.520	322615	82.46	1768



Gambar 1. Distribusi Variabel Numerik

Histogram pada Gambar 1 menunjukkan variabel numerik memiliki distribusi yang beragam. Beberapa variabel cenderung terpusat pada nilai tertentu (misalnya, variabel IPM), sementara yang lain lebih tersebar. Terdapat beberapa variabel numerik yang memiliki distribusi relatif normal (lonceng), seperti variabel IPM, APS, dan PDRB. Namun, banyak variabel lainnya yang cenderung miring atau tidak simetris, serta rentang nilai antar variabel cukup bervariasi.

Gambar 2 menunjukkan sebagian besar variabel memiliki *outlier*. *Outlier* teridentifikasi sebagai titik-titik data yang jauh dari kotak utama. Ini menunjukkan adanya nilai ekstrem. Rentang antarkuartil (IQR) setiap variabel cenderung berbeda yang mengindikasikan tingkat penyebaran data yang berbeda pada setiap variabel. *Outlier* akan dipertahankan pada analisis ini karena menjadi representasi nilai ekstrem yang dapat mewakili kondisi unik di suatu provinsi, serta dapat digunakan untuk membandingkan kondisi tertentu antar berbagai provinsi.



Gambar 2. Boxplot Variabel Numerik

Untuk identifikasi lebih lanjut mengenai perbedaan signifikan antara rata-rata kondisi akses pendidikan di pulau Jawa dan luar Jawa dilakukan uji *t* dua sampel independen. Data yang digunakan adalah rasio anak usia aktif SMA (16-18 tahun) dengan jumlah SMA tahun 2023.

Tabel 3. Hasil Uji Normalitas pada Data

Jenis Uji	Statistik Uji	<i>p-value</i>
Shapiro-Wilk	0.95549308	0.13512647

Berdasarkan Tabel 3, nilai statistik uji yang diperoleh adalah 0.955 dan *p-value* sebesar 0.135 dengan tingkat signifikansi 0.05. Ini mengindikasikan tidak ada bukti yang cukup untuk menolak hipotesis nol, bahwa data berdistribusi normal. Dengan demikian, data rasio anak usia aktif SMA (16-18 tahun) dengan jumlah SMA tahun 2023 berdistribusi normal. Hasil ini memungkinkan penggunaan analisis statistik parametrik, seperti uji *t* atau ANOVA.

Tabel 4. Hasil Perbedaan Mean Wilayah Jawa dan Luar Jawa

Jenis Uji	Statistik Uji	<i>p-value</i>
Uji <i>t</i> dua sampel independen	4.05982108	0.00025309

Hasil uji *t* dua sampel independen digunakan untuk mengetahui apakah terdapat perbedaan signifikan antara rata-rata dua kelompok yang tidak berhubungan, yaitu kelompok Jawa dan luar-Jawa. Pada Tabel 4, diperoleh nilai statistik uji sebesar 4.060 dan *p-value* sebesar 0.00025. Dengan tingkat signifikansi 0.05 diperoleh adanya perbedaan yang signifikan secara statistik antara rata-rata kedua kelompok tersebut. Artinya, terdapat perbedaan yang nyata antara kelompok Jawa dan luar Jawa dalam rasio anak usia aktif SMA (16-18 tahun) dengan jumlah SMA tahun 2023. Ini mengindikasikan adanya disparitas akses pendidikan jenjang SMA yang perlu dianalisis lebih lanjut.

Analisis lebih lanjut menggunakan regresi linier berganda dilakukan untuk mengidentifikasi faktor-faktor yang memengaruhi akses pendidikan di berbagai provinsi di Indonesia, khususnya dalam aspek Angka Partisipasi Sekolah (APS). Pembentukan model ini melibatkan beberapa representasi faktor sosial-ekonomi di tiap provinsi, seperti Produk Domestik Regional Bruto per Kapita atas Dasar Harga Berlaku (PDRB), Jumlah SMA, transformasi logaritma natural dari Persentase Kemiskinan, Indeks Ketimpangan Gender (IKG), Tingkat Pengangguran Terbuka (TPT), Indeks Pembangunan Manusia (IPM), Persentase Jumlah Guru SMA, dan Tingkat Penyelesaian Pendidikan menurut Jenjang Pendidikan SMA (TPP).

Tabel 5. Hasil Regresi Linier Berganda

Variabel Independen	Koefisien Regresi β (standard error)	t score	p-value	Keterangan
Intersep	-370.8838 (102.573)	-3.616	0.001	Signifikan
ln(Persentase Kemiskinan)	5.7071 (2.149)	2.656	0.014	Signifikan
ln(IKG)	0.5377 (1.573)	0.342	0.735	Tidak Signifikan
ln (TPT)	-0.8397 (3.736)	-0.225	0.824	Tidak Signifikan
PDRB	-8.3×10^{-6} (1.69×10^{-5})	-0.490	0.628	Tidak Signifikan
ln(IPM)	41.1479 (32.358)	1.276	0.214	Tidak Signifikan
Jumlah SMA	-0.0037 (0.004)	-0.888	0.383	Tidak Signifikan
ln(Persentase Jumlah Guru)	-0.9861 (1.877)	-0.525	0.604	Tidak Signifikan
ln(TPP)	58.3872 (20.885)	2.796	0.010	Signifikan

Dari hasil taksiran OLS diperoleh taksiran model sebagai berikut:

$$\begin{aligned} \text{APS} = & -370.8838 + 5.7071 \cdot \ln(\text{Persentase Kemiskinan}) + 0.5377 \cdot \ln(\text{IKG}) - 0.8397 \cdot \ln(\text{TPT}) \\ & - 8.3 \times 10^{-6} \cdot \text{PDRB} + 41.1479 \cdot \ln(\text{IPM}) - 0.0037 \cdot \text{Jumlah SMA} \\ & - 0.9861 \cdot \ln(\text{Persentase Jumlah Guru}) + 58.3872 \cdot \ln(\text{TPP}) \end{aligned}$$

Tabel 6. Nilai Kualitas Model Regresi Linier Berganda

R Squared	Adjusted R Squared	Statistik Uji F	p-value
0.593	0.462	4.546	0.00164

Berdasarkan Tabel 6, diketahui bahwa model tersebut memiliki *R Squared* sebesar 0.593. Artinya, model dapat menjelaskan 59.3% variasi dari Angka Partisipasi Sekolah (APS). Selanjutnya, diperoleh *Adjusted R Squared* sebesar 0.462 yang berarti setelah mempertimbangkan jumlah variabel, model mampu menjelaskan 46.2% variasi APS. Terakhir, dengan nilai statistik uji F sebesar 4.546 dan *p-value* sebesar 0.00164 menunjukkan bahwa model secara keseluruhan signifikan atau dengan kata lain setidaknya ada satu parameter pada model yang menjelaskan variabel APS secara signifikan.

Tabel 7. Hasil Uji Multikolinieritas		
Variabel Independen	Nilai VIF	Keterangan
(Konstanta)	18023.622005	-
ln(Persentase Penduduk Miskin)	1.790095	Tidak ada multikolinieritas
ln(IKG)	1.259060	Tidak ada multikolinieritas
ln (TPT)	1.868557	Tidak ada multikolinieritas
PDRB	1.826825	Tidak ada multikolinieritas
ln(IPM)	4.915085	Tidak ada multikolinieritas
Jumlah SMA	4.246380	Tidak ada multikolinieritas
ln(Persentase Jumlah Guru)	4.629653	Tidak ada multikolinieritas
ln(TPP)	3.131481	Tidak ada multikolinieritas

Selanjutnya, dilakukan uji multikolinieritas dengan melihat nilai VIF pada setiap variabel independen. Hasil VIF menunjukkan tidak ada masalah multikolinieritas yang signifikan. Seluruh nilai VIF variabel independen berada di bawah ambang batas 10, yang berarti hubungan antar variabel cukup rendah.

Uji asumsi pada model regresi linier berganda menunjukkan hasil yang memadai. Pada uji normalitas residual menggunakan Shapiro-Wilk, diperoleh nilai statistik uji sebesar 0.9566 dengan p-value 0.1938, yang berarti hipotesis nol gagal ditolak, sehingga residual model mengikuti distribusi normal. Selain itu, uji autokorelasi residual menggunakan Durbin-Watson menghasilkan nilai statistik uji sebesar 1.70, yang mendekati nilai 2, menunjukkan bahwa tidak terdapat autokorelasi yang signifikan dalam residual. Dengan demikian, model regresi linier berganda ini memenuhi asumsi normalitas dan tidak menunjukkan autokorelasi yang berarti.

Tabel 8. Hasil Uji Heterogenitas Spasial			
Jenis Uji	Statistik Uji	Derajat Kebebasan	p-value
Uji Breusch-Pagan	15.6136	8	0.0483

Hipotesis nol (H_0) dalam uji Breusch-Pagan menyatakan bahwa tidak terdapat heterogenitas spasial (artinya, tidak ada variasi sistematis yang tergantung pada lokasi spasial dalam residual). Sedangkan hipotesis alternatifnya (H_1) menunjukkan bahwa terdapat heterogenitas spasial (artinya, ada variasi sistematis dalam residual yang tergantung pada lokasi spasial). Dengan nilai p-value sebesar 0.0483 menunjukkan bahwa hipotesis nol ditolak. Artinya, heterogenitas spasial terdeteksi dalam model tersebut, atau dengan kata lain terdapat variasi residual yang sistematis yang dipengaruhi oleh faktor spasial (lokasi).

Selanjutnya, dilakukan pemodelan *Geographically Weighted Regression* (GWR). Pendugaan parameter akan menggunakan *Weighted Least Square* (WLS) yang melibatkan matriks pembobot berdasarkan jarak *Euclidean*, fungsi pembobot kernel, dan *bandwidth optimum*. Pada penelitian ini, akan dibentuk model *Geographically Weighted Regression* (GWR) dengan fungsi pembobot *Kernel Gaussian* dan *Kernel Bisquare* sehingga dapat dilakukan perbandingan untuk mendapatkan model

terbaik untuk data yang sedang diteliti. Perbandingan akan dilakukan berdasarkan nilai nilai koefisien determinasi (*R Squared*) dan *Akaike's Information Criterion* (AIC).

Tabel 9. Perbandingan Antar Model

Model	Bandwidth Optimum	R Squared	AIC
Regresi Global	-	0.46	207.628
GWR dengan <i>Kernel Gaussian</i>	11.25694	0.73	187.499
GWR dengan <i>Kernel Bisquare</i>	32.24075	0.69	190.1015

Berdasarkan Tabel 9 didapatkan bahwa nilai AIC dari model GWR dengan fungsi pembobot *Kernel Gaussian* memiliki nilai AIC yang lebih rendah di antara ketiga model yang diajukan. Selain itu, model tersebut juga memiliki nilai *R Squared* yang lebih tinggi dibandingkan model lainnya. Hal ini menunjukkan bahwa model GWR dengan fungsi pembobot *Kernel Gaussian* lebih sesuai dalam menjelaskan faktor-faktor yang memengaruhi kesenjangan akses pendidikan di Indonesia pada tahun 2023.

Selanjutnya, akan dilakukan uji parsial-*t* pada tiap parameter pada model GWR dengan fungsi pembobot *Kernel Gaussian*. Hal ini bertujuan untuk mengetahui apakah variabel-variabel independen berpengaruh signifikan terhadap variabel respon secara individual. Hipotesis uji parsial-*t* untuk model GWR adalah sebagai berikut:

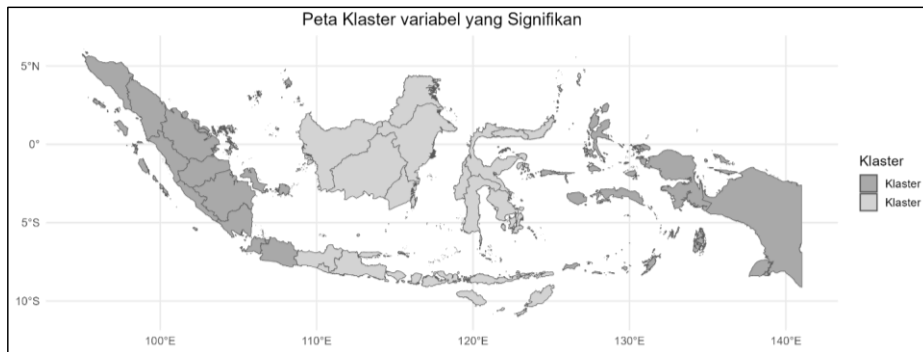
$$H_0: \beta_k(u_i, v_i) = 0, k = 1, 2, 3, \dots, 8; i = 1, 2, 3, \dots, 34$$

$$H_1: \text{minimal terdapat satu } \beta_k(u_i, v_i) \neq 0, k = 1, 2, 3, \dots, 8; i = 1, 2, 3, \dots, 34$$

Hasil dari pengujian parameter menunjukkan kelompok variabel independen yang memiliki pengaruh signifikan terhadap variabel respon. Kemudian akan dibentuk klaster berdasarkan variabel independen yang signifikan. Berdasarkan hasil analisis, terbentuk beberapa klaster yaitu:

Tabel 10. Klaster Model GWR Berdasarkan Variabel-Variabel Prediktor yang Signifikan

Klaster	Variabel yang Signifikan	Provinsi
1	TPP	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Maluku, Maluku Utara, Papua Barat, dan Papua.
2	Persentase Penduduk Miskin dan TPP	Jawa Tengah, Daerah Istimewa Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, dan Sulawesi Barat.



Gambar 3. Peta Persebaran Klaster Berdasarkan Variabel-Variabel yang Signifikan

Berdasarkan hasil klaster variabel yang signifikan dari model *Geographically Weighted Regression* (GWR), terdapat dua kelompok provinsi yang menunjukkan pola signifikan dalam kesenjangan akses pendidikan di Indonesia. Klaster pertama mencakup provinsi-provinsi seperti Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Maluku, Maluku Utara, Papua Barat, dan Papua, yang menunjukkan bahwa variabel tingkat penyelesaian pendidikan pada jenjang SMA berperan signifikan dalam memengaruhi akses pendidikan di wilayah tersebut. Hal ini menunjukkan bahwa di provinsi-provinsi tersebut, faktor pendidikan, khususnya tingkat kelulusan SMA, menjadi elemen yang mempengaruhi tingkat akses pendidikan, yang sejalan dengan temuan sebelumnya bahwa pendidikan yang lebih tinggi berperan penting dalam mendukung mobilitas sosial.

Sementara itu, klaster kedua mencakup provinsi-provinsi seperti Jawa Tengah, Daerah Istimewa Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, dan Sulawesi Barat, yang menunjukkan bahwa baik variabel tingkat penyelesaian pendidikan pada jenjang SMA maupun persentase penduduk miskin memiliki pengaruh signifikan terhadap kesenjangan akses pendidikan. Hal ini mengindikasikan bahwa di provinsi-provinsi ini, kombinasi antara tingkat kelulusan SMA dan kemiskinan berperan dalam membatasi akses masyarakat terhadap pendidikan, yang mendukung hasil penelitian sebelumnya bahwa ketimpangan sosial-ekonomi memengaruhi kualitas pendidikan.

Terdapat beberapa variabel lain yang lebih relevan dibandingkan dengan enam variabel yang tidak signifikan dalam mempengaruhi akses pendidikan di Indonesia. Salah satunya adalah anggaran pendidikan, yang berperan penting dalam meningkatkan kualitas pendidikan dan mengurangi ketimpangan antarprovinsi. Provinsi dengan pengeluaran pendidikan yang lebih

tinggi cenderung memiliki akses yang lebih baik dan fasilitas yang lebih lengkap (Muttaqin, 2018). Selain itu, ketimpangan pendapatan juga memainkan peran signifikan, di mana provinsi dengan tingkat ketimpangan pendapatan yang lebih tinggi sering kali mengalami kesulitan dalam memberikan pendidikan yang merata, karena perbedaan sosial-ekonomi yang tajam (Setyadi, 2022).

Faktor lain yang relevan adalah pendidikan orang tua, terutama pendidikan ibu, yang memengaruhi dukungan keluarga terhadap pendidikan anak. Keluarga dengan tingkat pendidikan lebih tinggi lebih cenderung memberikan dukungan materiil dan non-materiil untuk pendidikan anak mereka (Setyadi, 2022). Kemudian, tingkat urbanisasi juga memengaruhi akses pendidikan, di mana wilayah urban memiliki lebih banyak fasilitas pendidikan dan peluang yang lebih baik dibandingkan daerah pedesaan yang terbatas (Muttaqin, 2018). Meskipun tidak selalu signifikan di seluruh wilayah, ketimpangan gender tetap menjadi faktor penting yang menghambat akses pendidikan, terutama untuk perempuan di daerah-daerah konservatif (Setyadi, 2022).

Simpulan

Berdasarkan hasil analisis data, dapat disimpulkan bahwa terdapat disparitas atau kesenjangan akses pendidikan yang signifikan pada jenjang Sekolah Menengah Atas (SMA) antara wilayah Jawa dan luar Jawa. Hal ini dapat ditunjukkan dari hasil uji t dua sampel independen pada rasio anak usia 16-18 tahun (usia anak SMA) terhadap jumlah SMA pada tahun 2023. Ini mengindikasikan bahwa anak-anak di luar Jawa memiliki peluang lebih rendah untuk melanjutkan pendidikan ke tingkat SMA. Disparitas ini berpotensi menghambat pemerataan kualitas sumber daya manusia serta menghambat tercapainya Tujuan Pembangunan Berkelanjutan (TPB), yaitu Tujuan ke-4. Kedua, analisis *Geographically Weighted Regression* (GWR) memberikan pemahaman yang lebih mendalam terkait heterogenitas spasial dari faktor-faktor yang memengaruhi akses pendidikan di Indonesia. Berdasarkan analisis GWR, diperoleh dua kluster di mana pada kluster pertama tingkat penyelesaian pendidikan SMA (TPP) menjadi faktor yang dominan, dan kluster kedua baik TPP maupun persentase penduduk miskin secara bersama-sama memengaruhi akses pendidikan. Hal ini mengindikasikan bahwa kebijakan peningkatan akses pendidikan perlu disesuaikan dengan karakteristik tiap provinsi, mengingat perbedaan kondisi sosial dan ekonomi di masing-masing wilayah.

Namun, penelitian ini hanya berfokus pada tingkat SMA dan belum mempertimbangkan disparitas pada jenjang pendidikan lainnya. Penelitian lebih lanjut diperlukan untuk menganalisis disparitas akses pendidikan secara komprehensif, mulai dari tingkat dasar hingga perguruan tinggi. Selain itu, dari delapan faktor yang dianalisis, hanya dua faktor yang menunjukkan pengaruh signifikan terhadap akses pendidikan. Hal ini mengindikasikan adanya kompleksitas yang lebih

Commented [EO1]: cmiw

tinggi dalam dinamika akses pendidikan di Indonesia, yang memerlukan penelitian lebih lanjut untuk mengidentifikasi faktor-faktor lain yang mungkin berpengaruh signifikan.

Daftar Pustaka

- Chen, T., Yang, K., & Tu, J. (2019). Homoscedasticity: An overlooked critical assumption for linier regression. *General Psychiatry*, 32(1), e100080. <https://doi.org/10.1136/gpsych-2019-100080>
- Edo, A., Yasin, M. (2024). Dampak Kesenjangan Akses Pendidikan dan Faktor Ekonomi Keluarga terhadap Mobilitas Sosial. *SINOVA: Jurnal Ilmu Pendidikan dan Sosial*, 2 (3). <https://miftahul-ulum.or.id/ojs/index.php/jps/article/view/175>
- Katz, M. H. (2006). Assumptions of multiple linier regression, multiple logistic regression, and proportional hazards analysis. *Multivariable Analysis: A Practical Guide for Clinicians*, 5(1), 25-35. Cambridge University Press. <https://doi.org/10.1017/S0007114500132871>
- Khoiriyani, F., Dani, F. Z. D. P., Ansyah, R. H. A.. (2023). Kesenjangan Kualitas Layanan Madrasah Aliyah Negeri di Jawa dan Luar Jawa. *Jurnal SMART: Studi Masyarakat, Religi, dan Tardisi*, 9 (1). <https://doi.org/10.18784/smart.v9i1.1820>
<https://journal.blasemarang.id/index.php/smart/article/view/1820>
- Muttaqin, T. (2018). Determinants of Unequal Access to and Quality of Education in Indonesia. *Journal Perencanaan Pembangunan: The Indonesian Journal of Development Planning*, 2(1), 1-23. <https://journal.bappenas.go.id/index.php/jpp/article/view/27>
- Peña, E. A., & Slate, E. H. (2006). Global validation of linier model assumptions. *Journal of the American Statistical Association*, 101(476), 517-527. <https://doi.org/10.1198/016214505000000637>
- Setyadi, S. (2022). Inequality of Education in Indonesia by Gender, Socioeconomic Background. *EKO-REGION*, 17(1), 27-36
<https://media.neliti.com/media/publications/383462-inequality-of-education-in-indonesia-by-21fd9d66.pdf>
- Soliha, N. L., Widyaningsih, Y., Lestari, D. (Juni 2023). Analisis Faktor-Faktor yang Menjelaskan Kasus AIDS Provinsi Jawa Timur Menggunakan Model *Geographically Wiegthed Logistic Regression* (GWR). *Jurnal Statistika dan Aplikasinya*, 1 (1), 37-48. DOI: 10.21009/JSA.07104
https://drive.google.com/file/d/1_aHvnyZlkgQhUv-uKCW2tRwZKcOz8tQW/view?usp=sharing
- Tranmer, M., Murphy, J., Elliot, M., and Pampaka, M. (2020). *Multiple linier regression*. The Cathie Marsh Centre for Research on Statistics and the Social Sciences. <https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/multiple-linier-regression.pdf>

Lampiran

1. Hasil Uji Normalitas Residual

Jenis Uji	Statistik Uji	<i>p-value</i>
Shapiro-Wilk	0.9566	0.1938

2. Hasil Uji Autokorelasi Residual

Jenis Uji	Statistik Uji	Keterangan
Durbin-Watson	1.70	Tidak ada autokorelasi