

GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION UNTUK MENANGANI OVERDISPERSI PADA JUMLAH PENDUDUK MISKIN

Nova Delvia^{1*}, Mustafid², Hasbi Yasin³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*novadelvia511@gmail.com

ABSTRACT

Poverty is a condition that is often associated with needs, difficulties and deficiencies in various life circumstances. The number of poor people in Indonesia increase in 2020. This research focus on modelling the number of poor people in Indonesia using Geographically Weighted Negative Binomial Regression (GWNBR) method. The number of poor people is count data, so analysis used to model the count data is poisson regression. If there is overdispersion, it can be overcome using negative binomial regression. Meanwhile to see the spatial effect, we can use the Geographically Weighted Negative Binomial Regression method. GWNBR uses a adaptive bisquare kernel for weighting function. GWNBR is better at modelling the number of poor people because it has the smallest AIC value than poisson regression and negative binomial regression. While the GWNBR method obtained 13 groups of province based on significant variables.

Keywords: GWNBR, Poisson regression, Spatial Effect, Overdispersion, Poverty

1. PENDAHULUAN

Kemiskinan adalah permasalahan yang sifatnya multidimensional. Pendekatan dengan satu bidang ilmu tertentu tidaklah mencukupi untuk mengurai makna dan fenomena yang menyertainya. Definisi secara umum yang lazim dipakai dalam perhitungan dan kajian-kajian akademik adalah pengertian kemiskinan yang diperkenalkan oleh Bank Dunia yaitu sebagai ketidakmampuan mencapai standar hidup minimum (World Bank, 1990).

Penelitian oleh Brunson (1996) menunjukkan bahwa model regresi global tidak dapat menjelaskan hubungan antara beberapa variabel jika terdapat faktor geografis, akan tetapi GWR dapat menjelaskan hubungan yang berbeda pada setiap titik lokasi pengamatan. Penelitian pemodelan angka kematian bayi dengan pendekatan (GWPR) di Provinsi Jawa Timur oleh Ardiyanti (2010) menunjukkan bahwa model lokal atau model GWPR lebih baik dari pada model global atau model regresi Poisson.

Namun, GWPR gagal untuk memperhitungkan overdispersi yang ditemukan dalam pengamatan jumlah kasus kemiskinan di Provinsi Jambi. Sehingga dilakukan pendekatan dengan metode *Geographically Weighted Negative Binomial Regression* (GWNBR) untuk memungkinkan data diskrit dimodelkan dalam bentuk non-stasioner dan untuk menangani kasus overdispersi.

Jumlah penduduk miskin di Indonesia merupakan data count yang mengikuti distribusi Poisson. Dalam kasus Regresi Poisson sering dijumpai nilai varian lebih besar dari mean (overdispersi). Salah satu cara mengatasi kasus overdispersi dalam regresi Poisson adalah regresi Binomial Negatif. Perbedaan faktor-faktor yang berpengaruh di masing-masing wilayah menunjukkan adanya pengaruh kondisi lokal dari suatu wilayah. Oleh karena itu, pada penelitian ini dilakukan analisis dengan memperhatikan faktor spasial menggunakan metode *Geographically Weighted Negative Binomial Regression*.

2. TINJAUAN PUSTAKA

2.1. Kemiskinan

Secara umum kemiskinan adalah suatu permasalahan yang sering dikaitkan dengan ekonomi masyarakat, padahal jika dilihat secara luas kemiskinan dapat dilihat dari sudut pandang sosial maupun budaya dari masyarakat. Kemiskinan merupakan sebuah permasalahan yang sering dihadapi oleh masyarakat dimana terdapat kondisi ketidakmampuan untuk memenuhi kehidupan hidup sehari-hari dimulai dari pemenuhan papan, sandang, maupun pangan. Fenomena seperti hal ini biasa terjadi dikarenakan rendahnya penghasilan masyarakat dan juga rendahnya kualitas sumber daya manusia itu sendiri (Sajogyo, 1977). Pada tahun 2020 terjadi kenaikan jumlah penduduk miskin di Indonesia. Variabel yang digunakan antara lain, tingkat pengangguran terbuka tiap provinsi, kepadatan penduduk tiap provinsi, pertumbuhan ekonomi tiap provinsi, persentase rumah tangga yang menempati rumah dengan status tidak milik sendiri, dan rasio tenaga kesehatan tiap provinsi.

2.2. Multikolinearitas

Model regresi terbaik seharusnya tidak terjadi multikolinearitas, karena adanya korelasi antar variabel prediktor dapat menyebabkan estimasi parameter regresi yang dihasilkan akan memiliki error yang sangat besar. Pendeteksian kasus multikolinearitas menurut Hocking (1996) dapat dilihat dengan beberapa cara sebagai berikut :

- Jika koefisien korelasi Pearson (r_{ij}) antar variabel prediktor lebih besar dari 0,95 maka terdapat korelasi antar variabel tersebut
- Nilai VIF (*Variance Inflation Factor*) lebih besar dari 10 menunjukkan adanya multikolinearitas. Nilai VIF dinyatakan sebagai berikut :

$$VIF_j = \frac{1}{1 - R_j^2}$$

Dengan R_j^2 merupakan nilai koefisien determinasi antara variabel x_j dengan variabel prediktor lainnya.

2.3. Regresi Poisson

Regresi poisson adalah salah satu regresi yang dapat menggambarkan hubungan antara variabel respon (Y) dimana variabel respon berdistribusi poisson dengan variabel prediktor (X). dimana fungsi distribusinya adalah sebagai berikut (Myers, 1990) :

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots$$

dengan μ merupakan rata – rata variabel random Y yang berdistribusi poisson dimana nilai rata – rata dan varian variabel random Y memiliki nilai lebih dari nol.

Persamaan regresi poisson dapat dinyatakan sebagai berikut :

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

μ_i merupakan model regresi poisson dengan x_{ik} sebagai variabel prediktor dan β_k sebagai parameter regresi yang akan ditaksir.

Metode yang digunakan untuk menduga parameter model regresi poisson adalah metode *Maximum Likelihood Estimation* (MLE). Dalam regresi poisson, parameter yang diestimasi adalah β_k . Untuk mendapatkan nilai taksiran langkah yang dilakukan adalah membentuk fungsi *likelihood* dari fungsi poisson dan dilanjutkan dengan iterasi Newton Raphson.

Pengujian parameter model regresi poisson bertujuan untuk menguji apakah parameter berpengaruh signifikan terhadap variabel respon (Y). Uji signifikansi secara serentak menggunakan *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis sebagai berikut (McCullagh & Nelder, 1989):

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

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut :

$$D(\hat{\beta}) = -2 \ln \Lambda = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 (\ln L(\hat{\Omega}) - L(\hat{\omega}))$$

Dimana $D(\hat{\beta})$ adalah nilai devians model regresi poisson dan $L(\hat{\omega})$ merupakan suatu fungsi *likelihood* untuk model sederhana tanpa melibatkan variabel prediktor, sedangkan $L(\hat{\Omega})$ merupakan suatu fungsi *likelihood* untuk model yang lengkap yang melibatkan variabel prediktor. Jika nilai $D(\hat{\beta})$ semakin kecil maka semakin kecil pula tingkat kesalahan yang dihasilkan oleh model. Diketahui daerah penolakan adalah tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $\chi^2_{(p,\alpha)}$ yang artinya minimal terdapat satu parameter dalam regresi poisson yang berpengaruh signifikan terhadap variabel respon.

Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon dengan hipotesis sebagai berikut :

$$H_0 : \beta_k = 0, k = 1, 2, \dots, p$$

$$H_1 : \beta_k \neq 0$$

Statistik uji yang digunakan sebagai berikut :

$$Z_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)}$$

$se(\hat{\beta}_k)$ adalah standar error, didapatkan dari elemen diagonal ke-(k+1) dari $var(\hat{\beta})$ dengan $var(\hat{\beta}) = -E(H^{-1}(\hat{\beta}))$. Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon dalam model regresi poisson.

2.4. Overdispersi

Overdispersi dalam regresi poisson terjadi apabila nilai varians data lebih besar daripada nilai meannya. Overdispersi dapat dituliskan sebagai berikut (McCullagh & Nelder, 1989) :

$$Var(Y) > E(Y) \tag{2.12}$$

Overdispersi dapat dideteksi dengan nilai dispersi *Pearson chi Square* atau devians yang dibagi dengan derajat kebebasannya. Jika hasil pembagian kedua nilai tersebut lebih besar dari 1, maka dapat dikatakan terjadi overdispersi. Jika pada regresi poisson terjadi overdispersi, maka salah satu alternatif yang dapat digunakan adalah regresi binomial negatif (Hardin & Hilbe, 2007).

2.5. Regresi Binoial Negatif

Regresi binomial negatif merupakan salah satu solusi untuk mengatasi adanya kasus overdispersi. Model binomial negatif merupakan paduan antara distribusi poisson dan gamma. Dengan fungsi masa peluang binomial negatif :

$$f(y, \mu, \theta) = \frac{\Gamma(y + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) y!} \left(\frac{1}{1 + \theta\mu} \right)^{1/\theta} \left(\frac{\theta\mu}{1 + \theta\mu} \right)^y, y = 0, 1, 2, \dots$$

Bentuk umum dari regresi binomial negatif adalah :

$$y_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Dimana y adalah variabel respon berdistribusi binomial negatif yang berbentuk vektor dengan ukuran $(n \times 1)$, x adalah variabel prediktor yang berbentuk matriks dengan ukuran $(n \times (p+1))$ dan β adalah parameter yang berbentuk vektor dengan ukuran $((p+1) \times 1)$.

Estimasi parameter model regresi binomial negatif menggunakan metode maksimum likelihood dengan prosedur Newton Raphson.

Uji kesesuaian model regresi binomial negatif dengan serentak sebagai berikut :

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

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut :

$$D(\hat{\beta}) = -2 \ln \Lambda = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\omega}_0)} \right) = 2 (\ln L(\hat{\omega}) - \ln L(\hat{\omega}_0))$$

Kriteria penolakan adalah tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $\chi^2_{(p, \alpha)}$ yang artinya minimal terdapat satu parameter dalam regresi binomial negatif yang berpengaruh signifikan terhadap variabel respon.

Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon dengan hipotesis sebagai berikut :

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Statistik uji yang digunakan sebagai berikut :

$$Z_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)}$$

Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon dalam model regresi binomial negatif.

2.6. Pengujian Spasial

Regresi spasial merupakan salah satu metode yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor dengan memperhatikan aspek lokasi atau spasial. Pengujian spasial dibagi menjadi dua yaitu pengujian heterogenitas spasial dan pengujian dependensi spasial.

Pengujian heterogenitas spasial ini dilakukan untuk melihat perbedaan karakteristik antara satu titik pengamatan dengan titik pengamatan lainnya menyebabkan adanya heterogenitas spasial. Untuk melihat adanya heterogenitas spasial pada data dapat dilakukan pengujian *Breusch – Pagan* (Anselin, 1988). Hipotesis yang digunakan adalah sebagai berikut :

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

$$H_1 : \text{paling sedikit ada satu } \sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n \text{ (varians antar lokasi berbeda)}$$

Statistik uji yang digunakan adalah sebagai berikut :

$$BP = \left(\frac{1}{2} \right) f^T Z (Z^T Z)^{-1} Z^T f \sim \chi^2_{(p)}$$

Kriteria penolakan yaitu tolak H_0 jika statistik uji $BP > \chi^2_{(\alpha, p)}$ yang artinya adalah variansi antar lokasi berbeda.

Pengujian dependensi spasial dilakukan untuk melihat apakah pengamatan di suatu lokasi berpengaruh terhadap pengamatan di lokasi lain yang letaknya berdekatan. Statistik uji yang sering digunakan dalam autokorelasi spasial adalah Moran's I. Hipotesis yang digunakan dalam pengujian dependensi spasial adalah sebagai berikut:

$H_0: I = 0$ (tidak ada dependensi spasial)

$H_1: I \neq 0$ (terdapat dependensi spasial)

Statistik uji yang digunakan adalah sebagai berikut:

$$Z_{I \text{ hit}} = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{Var}(\hat{I})}}$$

Dengan

$$\hat{I} = \frac{n \sum_{i=1}^n \sum_{j=i}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=i}^n w_{ij}) \sum_{i=1}^n (y_i - \bar{y})^2}$$

Kriteria penolakan yaitu tolak H_0 jika nilai $|Z_{I \text{ hit}}| > Z_{\alpha/2}$ yang artinya terdapat dependensi spasial.

2.7. Bandwidth dan Pembobot Optimum

Keragaman spasial yang terjadi pada suatu pengamatan diperlukan matriks pembobot. Fungsi kernel adaptif yaitu fungsi kernel yang memiliki *bandwidth* yang berbeda pada setiap lokasi pengamatan. Fungsi pembobot yang digunakan adalah fungsi *adaptive bisquare* kernel sebagai berikut :

$$W_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i}\right)^2\right)^2, & \text{untuk } d_{ij} \leq h_i \\ 0, & \text{untuk } d_{ij} > h_i \end{cases}$$

Dengan :

$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ adalah jarak *euclidean* antara lokasi ke-i dan lokasi ke-j
 $h_i = \text{adaptive bandwidth}$

Pemilihan *bandwidth* optimum menjadi sangat penting karena akan memengaruhi ketepatan model terhadap data, yaitu mengatur varians dan bias model. Penentuan *bandwidth* optimum dilakukan menggunakan metode *Cross Validation* (CV), yang dirumuskan sebagai berikut :

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h_i))^2$$

$\hat{y}_{\neq i}(h_i)$ merupakan nilai penaksir y_i dengan pengamatan lokasi (u_i, v_i) dihilangkan dari proses penaksiran.

2.8. Geographically Weighted Negative Binomial Regression

Model GWNBR akan menghasilkan penaksir parameter lokal dengan masing – masing lokasi akan memiliki parameter yang berbeda – beda. Model GWNBR dapat dirumuskan sebagai berikut (Ricardo & Carvalho, 2013) :

$$y_i \sim NB \left[\exp \left(\sum_{k=0}^p \beta_k(u_i, v_i) x_{ik} \right), \theta(u_i, v_i) \right], i = 1, 2, 3, \dots, n$$

Dimana :

y_i = nilai observasi respon ke-i

x_{ik} = nilai observasi variabel prediktor ke-k pada pengamatan lokasi (u_i, v_i)

$\beta_k(u_i, v_i)$ = koefisien regresi variabel prediktor ke-k untuk setiap lokasi (u_i, v_i)

$\theta(u_i, v_i)$ = parameter dispersi untuk setiap lokasi (u_i, v_i)

Estimasi parameter model GWNBR dilakukan dengan menggunakan metode *Maximum Likelihood Estimation*. Fungsi likelihood dapat dituliskan sebagai berikut (Ricardo & Carvalho, 2013) :

$$L(\beta(u_i, v_i), \theta_i | y_i, x_i) = \prod_{i=1}^n \left(\prod_{r=0}^{y-1} \left(r + \frac{1}{\theta_i} \right) \right) \left(\frac{1}{y!} \right) \left(\frac{1}{1 + \theta_i \mu_i} \right)^{\frac{1}{\theta_i}} \left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right)^{\frac{1}{y_i}}$$

Pengujian parameter secara serentak dapat menggunakan *Maximum Likelihood Ratio Test* (MLRT). Dengan hipotesis sebagai berikut :

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k(u_i, v_i) \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut :

$$D(\hat{\beta}) = -2 \ln \Lambda = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 (\ln L(\hat{\Omega}) - L(\hat{\omega}))$$

Kriteria penolakan adalah tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $\chi^2_{(p, \alpha)}$ yang artinya minimal terdapat satu parameter dalam model GWNBR yang berpengaruh signifikan terhadap variabel respon.

Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dengan hipotesis sebagai berikut :

$$H_0 : \beta_k(u_i, v_i) = 0$$

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

Statistik uji yang digunakan sebagai berikut :

$$Z_{hit} = \frac{\hat{\beta}_k(u_i, v_i)}{se(\hat{\beta}_k(u_i, v_i))}$$

Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dalam model GWNBR.

3. METODE PENELITIAN

3.1. Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini adalah data kuantitatif dengan jumlah data berupa data setiap provinsi di Indonesia (34 provinsi). Data diambil dari Survei Sosial Ekonomi Nasional (SUSENAS) periode Maret 2020. Data tersebut dapat diakses dari web Badan Pusat Statistik (BPS). Variabel penelitian yang digunakan adalah:

- Jumlah penduduk miskin tiap provinsi di Indonesia tahun 2020 (Y)
- Tingkat pengangguran terbuka tiap provinsi (X_2)
- Kepadatan penduduk tiap provinsi (X_3)
- Pertumbuhan ekonomi tiap provinsi (X_4)
- Persentase rumah tangga yang menempati rumah dengan status tidak milik sendiri (X_5)
- Rasio tenaga kesehatan tiap provinsi (X_6)

3.2. Langkah-langkah Analisis

Berikut merupakan langkah – langkah untuk melakukan analisis faktor – faktor yang memengaruhi kenaikan jumlah penduduk miskin di Indonesia :

- Mendeskripsikan karakteristik jumlah penduduk miskin dan faktor – faktor yang memengaruhinya di Indonesia pada tahun 2020 menggunakan pemetaan wilayah untuk masing – masing variabel

- b. Pengujian kasus multikolinearitas berdasarkan kriteria koefisien korelasi dan VIF
- c. Menganalisis model regresi poisson dengan tiga langkah berikut
 - Penaksiran parameter model regresi poisson menggunakan metode MLE
 - Menguji signifikansi parameter model regresi poisson secara serentak dan parsial
 - Melakukan uji dispersi model regresi poisson
- d. Menganalisis model regresi binomial negatif dengan langkah sebagai berikut
 - Penaksiran parameter model regresi binomial negatif dengan metode MLE
 - Menguji signifikansi parameter model binomial negatif
- e. Memodelkan GWNBR untuk kasus kenaikan jumlah kemiskinan di Indonesia pada tahun 2020, dengan langkah sebagai berikut
 - Uji *Breusch – Pagan* untuk melihat heterogenitas spasial data dan uji Moran I untuk menguji dependensi spasial data
 - Menghitung jarak *Euclidean* antar lokasi pengamatan berdasarkan posisi geografis
 - Mendapatkan *bandwidth* optimal untuk setiap lokasi pengamatan dengan menggunakan *Cross Validation (CV)*
 - Menghitung matriks pembobot dengan menggunakan fungsi kernel *adaptive bisquare kernel*
 - Melakukan interpretasi model GWNBR yang didapatkan
 - Pemilihan model terbaik

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Sampel

Statistika deskriptif data jumlah penduduk miskin di Indonesia dan faktor – faktor yang diduga memengaruhinya berdasarkan perhitungan Program R. Output komputasi dari Program R diberikan pada Tabel 1.

Tabel 1. Statistika Deskriptif Data Jumlah Penduduk Miskin dan Faktor-Faktor yang Diduga Memengaruhi

Variabel	Mean	Median	Minimum	Maximum
Y	810285	367750	52700	4585970
X1	0,06034	0,05570	0,03320	0,10950
X2	742	103.5	9	15900
X3	0,04092	0,03540	0,00830	0,09240
X4	0,10069	0,07025	0,02160	0,36360
X5	0,047033	0,044466	0,003717	0,090909

Tabel 1 menunjukkan bahwa rata-rata jumlah penduduk miskin di Indonesia pada tahun 2020 cukup tinggi yaitu sebesar 810285 dengan jumlah penduduk miskin tertinggi yaitu 4585970 dan jumlah terendah yaitu 52700.

4.2. Pengujian Multikolinearitas

Ada beberapa cara untuk mendeteksi adanya kasus multikolinearitas, yaitu dengan melihat nilai koefisien korelasi pearson (r_{ij}) dan nilai *Variance Inflation Factor (VIF)*. Koefisien korelasi antar variabel prediktor dihitung menggunakan Program R. Output Program R diberikan pada Tabel 2.

Tabel 2. Koefisien Korelasi Antar Variabel Prediktor

	X1	X2	X3	X4	X5
X1	1,00000				
X2	0,49092	1,00000			
X3	0,18275	0,38411	1,00000		
X4	0,58305	0,66703	0,20684	1,00000	
X5	0,11596	-0,24066	-0,21260	-0,39992	1,00000

Tabel 2 menunjukkan bahwa semua variabel prediktor memiliki nilai koefisien korelasi pearson yang kurang dari 0,95 yang artinya tidak terdapat kasus multikolinearitas. *Variance Inflation Factor* (VIF) masing - masing variabel prediktor dihitung menggunakan rumus $VIF_j = \frac{1}{1-R_j^2}$ dengan Program R. Output Program R diberikan pada Tabel 3.

Tabel 3. VIF Masing-Masing Variabel Prediktor

Variabel	VIF
X1	2,027848
X2	2,082287
X3	1,230053
X4	2,917639
X5	1,594355

Tabel 3 menunjukkan bahwa nilai VIF dari masing-masing variabel prediktor memiliki nilai VIF kurang dari 10, yang artinya tidak terdapat kasus multikolinearitas. Sehingga dapat dilanjutkan ke pemodelan regresi poisson dan binomial negatif.

4.3. Analisis Model Regresi Poisson

Setelah dilakukan pemeriksaan kasus multikolinearitas antara variabel prediktor dilanjutkan pada pemodelan regresi poisson. Data jumlah penduduk miskin diasumsikan berdistribusi poisson karena merupakan data diskrit (*count*). Berikut merupakan estimasi parameter model regresi poisson :

Tabel 4. Estimasi Parameter Model Regresi Poisson

	Estimate	Std. Error	Z Value	P Value
Intercept	12,18	0,00089	13532,2	2×10^{-16}
X1	23,65	0,01393	1697,9	2×10^{-16}
X2	0,00017	0,00000013	1352,3	2×10^{-16}
X3	10,14	0,01075	942,9	2×10^{-16}
X4	-13,35	0,00609	-2193,6	2×10^{-16}
X5	10,66	0,01252	851,1	2×10^{-16}
Deviance = 20610617		DF = 28		
AIC = 20611132				

Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5% didapatkan nilai deviance ($D(\hat{\beta})$) sebesar 20611137. Nilai $\chi^2_{(5,0.05)}$ sebesar 11,07 lebih kecil dari nilai deviance ($D(\hat{\beta})$) sebesar 20611137 sehingga H_0 ditolak yang artinya paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Untuk pengujian secara parsial dengan taraf signifikansi 5% didapatkan nilai $|Z_{hit}|$ lebih besar dari nilai $Z_{\frac{(0.05)}{2}}$ sebesar 1,96 yang artinya semua variabel prediktor dalam model secara individu memberikan pengaruh yang signifikan terhadap jumlah penduduk miskin di Indonesia. Berdasarkan Tabel 4 didapatkan model regresi poisson sebagai berikut :

$$\ln(\hat{\mu}) = 12,18 + 23,65X_1 + 0,00017X_2 + 10,14X_3 - 13,35X_4 + 10,66X_5.$$

4.4. Pengujian Overdispersi

Data Regresi poisson memiliki asumsi bahwa nilai mean sama dengan varian (equidispersi) namun pada data jumlah penduduk ini diduga terjadi kasus overdispersi $Var(Y) > E(Y)$. Untuk mendeteksi kasus overdispersi digunakan nilai dispersi *Pearson chi Square* atau devians yang dibagi dengan derajat bebasnya. Perhitungan overdispersi menggunakan Program R didapatkan nilai dispersi sebesar 706226,7 nilai tersebut lebih besar dari 1 yang artinya data jumlah penduduk mengalami kasus overdispersi.

4.5. Analisis Model Regresi Binomial Negatif

Estimasi parameter model regresi binomial negatif menggunakan metode *Maximum Likelihood Estimation* dengan prosedur Newton Raphson.

Tabel 5 Estimasi Parameter Model Regresi Binomial Negatif

	Estimate	Std. Error	Z Value	P Value
Intercept	11,40	0,6724	16,957	2×10^{-16}
X1	19,89	9,977	1,993	0,04624
X2	0,000072	0,000075	0,958	0,33802
X3	19,01	7,253	2,621	0,00876
X4	-8,496	3,377	-2,516	0,01187
X5	16,01	8,960	1,787	0,07398
Deviance = 37,583			DF = 28	
AIC = 985,22				

Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5% didapatkan nilai deviance ($D(\hat{\beta})$) sebesar 37,583. Nilai $\chi^2_{(5,0.05)}$ sebesar 11,07 lebih kecil dari nilai deviance ($D(\hat{\beta})$) sebesar 37,583 sehingga H_0 ditolak yang artinya paling sedikit ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Untuk pengujian secara parsial dengan taraf signifikansi 5% didapatkan empat variabel dengan nilai $|Z_{hit}|$ lebih besar dari nilai $Z_{\frac{(0.05)}{2}}$ sebesar 1,96. Artinya hanya empat variabel prediktor dalam model secara individu memberikan pengaruh yang signifikan terhadap jumlah penduduk miskin di Indonesia. Berdasarkan Tabel 5 didapatkan model regresi poisson sebagai berikut :

$$\ln(\hat{\mu}) = 11,40 + 19,89X_1 + 0,000072X_2 + 19,01X_3 - 8,496X_4 + 16,01X_5.$$

4.6. Pengujian Aspek Spasial

Berdasarkan hasil pengujian heterogenitas diperoleh nilai statistik uji *Breusch-Pagan* sebesar 7,0091 dengan p-value 0,22. Dengan jumlah parameter 5 dan taraf signifikansi sebesar 5% didapatkan $\chi^2_{(5,0.05)}$ sebesar 11,07. Karena nilai $BP < \chi^2_{(5,0.05)}$ maka diambil keputusan gagal tolak H_0 yang berarti varians antar lokasi sama atau tidak terdapat perbedaan antar titik pengamatan satu dan titik pengamatan lainnya. Berdasarkan hasil pengujian dependensi spasial diperoleh statistik uji Moran's I sebesar 0,03748102 dengan nilai $E(\hat{I})$ sebesar -0,03030303 dan $\sqrt{Var(\hat{I})}$ sebesar 0,03443856.

$$Z_{I \text{ hit}} = \frac{0,03748102 + 0,03030303}{0,03443856} = 1,968$$

Dengan taraf signifikansi 5% maka didapatkan nilai $Z_{\frac{(0.05)}{2}}$ sebesar 1,96. Karena $Z_{I \text{ hit}} > Z_{\frac{(0.05)}{2}}$ atau p-value (0,04904) $< 0,05$ maka diambil keputusan tolak H_0 sehingga didapatkan

kesimpulan bahwa terdapat dependensi spasial yang berarti pengamatan suatu lokasi berpengaruh terhadap lokasi lain yang letaknya berdekatan.

Pengujian heterogenitas spasial dan dependensi spasial menunjukkan bahwa data jumlah penduduk miskin di Indonesia hanya memenuhi salah satu asumsi yaitu asumsi dependensi spasial. Walaupun demikian pada penelitian ini ingin melihat efek spasial sehingga dapat dilanjutkan pemodelan dengan menggunakan metode GWNBR.

4.7. Pemodelan GWNBR

Terdapat tiga macam pengujian parameter untuk pemodelan GWNBR yaitu, pengujian kesamaan model GWNBR dengan binomial negatif, pengujian serentak dan pengujian parsial.

a. Pengujian kesamaan model GWNBR dan binomial negatif

Berdasarkan hasil perhitungan didapatkan nilai F_{hit} sebesar 0,00218. Dengan taraf signifikansi 5% didapatkan $F_{(0.05,28,28)}$ sebesar 1,882 yang artinya bahwa tidak terdapat perbedaan yang signifikan antara model binomial negatif dengan model GWNBR. Namun dalam penelitian ini dipilih model GWNBR sehingga dilanjutkan untuk pengujian parameter.

b. Pengujian serentak

Berdasarkan hasil perhitungan menggunakan Program R dan didapatkan nilai devians model GWNBR sebesar 125503,3. Dengan taraf signifikansi 5% didapatkan $\chi^2_{(5,0.05)}$ sebesar 11,07 yang artinya bahwa paling tidak ada satu parameter model GWNBR yang signifikan berpengaruh.

c. Pengujian parsial

Berdasarkan hasil pengujian secara parsial diperoleh parameter yang signifikan berbeda – beda untuk tiap provinsi. Nilai $Z_{I hit}$ parameter tiap provinsi dibandingkan dengan $Z_{\frac{(0.05)}{2}}$, jika nilai $Z_{I hit} > Z_{\frac{(0.05)}{2}}$ (1,96) maka tolak H_0 sehingga variabel tersebut memberikan pengaruh signifikan terhadap model. Berikut ini merupakan tabel pengelompokan provinsi berdasarkan variabel yang signifikan.

Tabel 6. Pengelompokan Provinsi Berdasarkan Variabel yang Signifikan

Kel	Provinsi	Variabel yang signifikan
1	Sumatera Selatan,Lampung, Kep. Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tenggara, Sulawesi Barat, Papua Barat	X_1, X_2, X_3, X_4, X_5
2	Aceh	X_1, X_2, X_3, X_4
3	Jambi, Bengkulu	X_1, X_2, X_3, X_5
4	Nusa Tenggara Timur, Gorontalo	X_2, X_3, X_4, X_5
5	Sulawesi Tengah, Maluku Utara	X_1, X_3, X_4, X_5
6	Sulawesi Utara	X_1, X_2, X_3
7	Sumatera Barat, Riau	X_2, X_3, X_5
8	Kalimantan Utara, Papua	X_1, X_2
9	Sumatera Utara, Sulawesi Selatan	X_2, X_3
10	Kalimantan Timur	X_2, X_4
11	Kep. Riau	X_3, X_5
12	Nusa Tenggara Barat	X_2
13	Sulawesi Barat	X_3

Sebagai contoh akan disajikan pengujian parameter pada lokasi penelitian yang ke-23 (u_{23}, v_{23}) yaitu Provinsi Kalimantan Timur.

Tabel 7. Pengujian Parameter Model GWNBR di Provinsi Kalimantan Timur

Parameter	Estimasi	Z hitung
Intercept	1,6359	703991,116
X_1	0,0804	1,4329
X_2	0,00000010	8,0895
X_3	0,0614	1,8749
X_4	0,1128	9,3722
X_5	0,0914	1,2601

Suatu variabel memberikan pengaruh yang signifikan jika $|Z_{hit}| > Z_{\alpha/2}$, dengan taraf signifikansi 5% didapatkan nilai $Z_{(\frac{0.05}{2})}$ sebesar 1,96 sehingga dapat diketahui variabel yang signifikan adalah variabel X_2 , dan X_4 . Sehingga model GWNBR yang dapat dibentuk adalah sebagai berikut :

$$\ln(\hat{\mu}_{23}) = 1,6359 + 0,0804X_1 + 0,00000010X_2 + 0,0614X_3 + 0,1128X_4 + 0,0914X_5$$

Berdasarkan variabel yang signifikan dari model yang terbentuk di Kalimantan Timur dapat disimpulkan bahwa setiap pertambahan 1 jiwa/ km^2 (X_2) maka akan menambah jumlah kasus penduduk miskin sebesar $\exp(0,00000010) = 1,00000010 \approx 1$ kasus dengan asumsi variabel lain konstan. Setiap pertambahan 1 persen rumah tangga yang menempati rumah dengan status tidak milik sendiri (X_4) akan menambah jumlah kasus penduduk miskin sebesar $\exp(0,1128) = 1,1194 \approx 1$ kasus dengan asumsi variabel lain konstan. Hal ini sesuai karena dengan meningkatnya kepadatan penduduk dan rumah tangga yang menempati rumah dengan status tidak milik sendiri maka jumlah penduduk miskin juga akan naik.

4.8. Pemilihan Model Terbaik

Pemilihan model terbaik berdasarkan kriteria AIC pada model regresi poisson, regresi binomial negatif, dan GWNBR adalah sebagai berikut :

Tabel 8. Pemilihan Model Terbaik

Model Regresi	AIC
Regresi Poisson	20611132
Regresi Binomial Negatif	985,22
GWNBR	974,2981

Tabel 8 menunjukkan bahwa dari ketiga model tersebut, GWNBR memiliki nilai AIC paling kecil dibandingkan dengan regresi poisson dan regresi binomial negatif, sehingga GWNBR lebih baik dalam memodelkan jumlah kasus penduduk miskin di Indonesia pada tahun 2020.

5. KESIMPULAN

Pada tahun 2020 persebaran penduduk miskin di Indonesia terjadi di seluruh provinsi. Pemodelan jumlah penduduk miskin di Indonesia menggunakan regresi poisson didapatkan bahwa semua variabel berpengaruh signifikan terhadap model. Sedangkan dengan menggunakan regresi binomial negatif hanya diperoleh tiga variabel yang berpengaruh yaitu variabel X_1 , X_3 , dan X_4 . Berdasarkan hasil pemodelan GWNBR dengan fungsi pembobot kernel *adaptive bisquare* didapatkan pengelompokkan sebanyak 13 kelompok berdasarkan variabel-variabel yang signifikan. Pemodelan menggunakan metode GWNBR diperoleh nilai AIC terkecil sehingga dapat dikatakan metode ini paling tepat untuk memodelkan jumlah kasus penduduk miskin di Indonesia.

DAFTAR PUSTAKA

- Ardiyanti, S. T., & Purhadi. (2010). *Pemodelan Angka Kematian Bayi dengan Pendekatan Geographically Weighted Poisson Regression (GWPR) di Provinsi Jawa Timur*. Undergraduate Theses.
- Brundson, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationary. *Geographical Analysis*. 28 (4) , 281 - 298.
- Hardin, J. W., & Hilbe, J. M. (2007). *Generalized Linear Models and Extensions Second Edition*. Texas: Stata Press.
- Hocking, R. R. (1996). *Methods and Applications of Linear Models : Regression and The Analysis of Variance*. New York: John Wiley and Sons.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Ricardo, A., & Carvalho, T. (2013). *Geographically Weighted Negative Binomial Regression-Incorporating Overdispersion*. Business Media New York: Springer Science.
- Sajogyo, T. (1997). *Garis Kemiskinan dan Kebutuhan Minimum Pangan*. Bogor: LPSBIPB.
- WorldBank. (1990). *World Development Report 1990 : Poverty*. New York: Oxford University Press.