

Evaluasi Model Regresi GWR dan NBR untuk Analisis Kriminalitas di Banten: Tantangan Sampel Terbatas

Puspa Nur Afifah^{1*}, Sekar Ayu Permata², Wahyu Arjuna Saputra³

^{1,2,3}Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Jl. Syekh Nawawi Al-Bantani No 30, Serang, Indonesia

E-mail: ¹puspa.nuraffah@uinbanten.ac.id, ²20153025@uinbanten.ac.id, ³201530002@uinbanten.ac.id

*Corresponding Author

ABSTRACT

This study evaluates three regression models in analyzing crime rates in Banten Province, namely Poisson Regression, Geographically Weighted Regression (GWR), and Negative Binomial Regression (NBR). The data covers eight administrative regions with explanatory variables such as population density, population-to-police ratio, open unemployment rate, poverty rate, average years of schooling, and the Human Development Index. The results indicate that 1) the Poisson model suffers from severe overdispersion and fails to capture the influence of predictor variables, 2) the GWR model has the lowest AIC value, but produces identical parameters across all regions, indicating overfitting due to the limited number of regions, 3) the NBR model provides low deviation and competitive AIC, indicating more stable statistical performance on limited data. This study highlights the importance of considering sample size in spatial model selection and suggests the use of NBR as a viable alternative for overdispersive count data with a small number of units.

Keywords: Banten, GWR, Crime, Negative Binomial Regression, Spatial

ABSTRAK

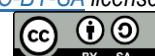
Penelitian ini mengevaluasi tiga model regresi dalam menganalisis tingkat kriminalitas di Provinsi Banten, yaitu Regresi Poisson, Regresi Geographically Weighted Regression (GWR), dan Regresi Binomial Negatif (NBR). Data mencakup delapan wilayah administratif dengan variabel penjelas seperti kepadatan penduduk, rasio penduduk terhadap polisi, tingkat pengangguran terbuka, angka kemiskinan, rata-rata lama sekolah, dan Indeks Pembangunan Manusia. Hasil menunjukkan bahwa 1) model Poisson mengalami overdispersi berat dan gagal menangkap pengaruh variabel prediktor, 2) model GWR memiliki nilai AIC terendah, namun menghasilkan parameter yang identik di semua wilayah, mengindikasikan overfitting akibat keterbatasan jumlah wilayah, 3) Model NBR memberikan deviasi rendah dan AIC yang kompetitif, menunjukkan kinerja statistik yang lebih stabil pada data terbatas. Studi ini menyoroti pentingnya mempertimbangkan ukuran sampel dalam pemilihan model spasial, serta menyarankan penggunaan NBR sebagai alternatif yang layak untuk data count yang overdispersi dengan jumlah unit kecil.

Kata kunci: Banten, GWR, Kriminalitas, Regresi Binomial Negatif, Spasial

Dikirim: Juni 2025; Diterima: Juli 2025; Dipublikasikan: September 2025

Cara sitasi: Afifah, P.N., Permata, S.A., & Saputra, W.A. (2025). Evaluasi Model Regresi GWR dan NBR untuk Analisis Kriminalitas di Banten: Tantangan Sampel Terbatas. *Teorema: Teori dan Riset Matematika*, 10(02), 171-180. DOI: <https://dx.doi.org/10.25157/teorema.v10i1.19317>.

This is an open access article under the [CC-BY-SA](#) license



PENDAHULUAN

Kriminalitas merupakan fenomena sosial yang tidak hanya dipengaruhi oleh perilaku individu, tetapi juga oleh kondisi lingkungan dan karakteristik wilayah tempat tinggal. Faktor-faktor seperti kepadatan penduduk, tingkat pengangguran, kemiskinan, dan akses terhadap pendidikan telah lama dikaji sebagai penentu utama tingkat kriminalitas suatu daerah (Eck, John & Weisburd, 2015). Namun demikian, penyebaran kejahatan tidak bersifat merata secara geografis. Berbagai penelitian menemukan bahwa kejahatan cenderung terkonsentrasi di lokasi-lokasi tertentu, seperti area padat penduduk, kawasan miskin, atau wilayah dengan layanan publik yang terbatas (Anselin, Cohen, Cook, Gorr, & Tita, 2000; Breetzke, 2018).

Di Indonesia sendiri, kajian spasial mengenai distribusi kriminalitas masih relatif terbatas, khususnya pada tingkat wilayah provinsi atau kabupaten/kota (Anozi & Novianda, 2023; Rahman & Prasetyo, 2018). Sebagian besar studi hanya memaparkan tren kriminalitas secara agregat tanpa mempertimbangkan ketimpangan spasial yang mungkin terjadi antarwilayah. Padahal, variabel-variabel sosial ekonomi seperti kemiskinan dan pendidikan dapat memiliki pengaruh yang berbeda-beda di tiap wilayah (Sugiharti *et al.*, 2022).

Salah satu provinsi yang mengalami peningkatan kasus kriminalitas secara signifikan adalah Provinsi Banten. Berdasarkan data Badan Pusat Statistik (BPS), jumlah kasus kriminal yang dilaporkan meningkat dari 5.118 kasus pada tahun 2022 menjadi 7.392 kasus pada tahun 2023 (BPS, 2024). Lonjakan ini menimbulkan pertanyaan mengenai faktor-faktor yang memengaruhi angka kriminalitas di wilayah ini serta bagaimana distribusinya secara spasial.

Model regresi spasial seperti *Geographically Weighted Regression* (GWR) banyak digunakan untuk menangkap variasi spasial dalam hubungan antara variabel prediktor dan variabel respons (Brunsdon *et al.*, 1996; Páez *et al.*, 2011). GWR memungkinkan estimasi parameter regresi yang berbeda untuk setiap lokasi, sehingga cocok untuk menganalisis fenomena yang tidak stasioner secara spasial seperti kriminalitas. Namun demikian, salah satu keterbatasan GWR adalah sensitivitasnya terhadap ukuran sampel yang kecil, yang dapat menyebabkan estimasi yang tidak stabil atau bahkan *overfitting* (Devkota *et al.*, 2014).

Sebagai alternatif, model Regresi Binomial Negatif (*Negative Binomial Regression*/NBR) dapat digunakan untuk mengatasi permasalahan overdispersi pada data cacah, khususnya ketika varians data melebihi rata-ratanya (Hilbe, 2011). NBR tidak mempertimbangkan struktur spasial secara eksplisit, tetapi lebih stabil ketika digunakan pada sampel kecil atau data dengan sebaran yang tidak homogen. Berbeda dengan penelitian sebelumnya yang menggunakan GWR pada data spasial yang luas, penelitian ini secara eksplisit mengevaluasi performa GWR dan NBR dalam konteks jumlah unit wilayah yang terbatas, yang masih jarang dikaji dalam konteks Indonesia, khususnya Provinsi Banten. Fokus utama terletak pada sejauh mana keterbatasan jumlah wilayah ($n = 8$) memengaruhi hasil analisis, serta model mana yang paling sesuai untuk konteks data yang tersedia. Hasil kajian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metodologi statistik spasial dalam konteks data terbatas, serta menjadi masukan bagi pengambil kebijakan dalam merancang intervensi berbasis wilayah. Penelitian ini bertujuan untuk: 1) Menganalisis faktor-faktor yang memengaruhi tingkat kriminalitas di Provinsi Banten tahun 2023. 2) Mendeteksi adanya overdispersi pada model Poisson. 3) Membandingkan performa GWR dan Regresi Binomial Negatif pada jumlah unit wilayah terbatas. 4) Menentukan model terbaik berdasarkan kriteria informasi statistik.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain studi eksploratif-komparatif. Tujuan utamanya adalah mengevaluasi kinerja beberapa model regresi dalam menjelaskan tingkat kriminalitas di Provinsi Banten berdasarkan data spasial terbatas. Data yang digunakan merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) dan lembaga pemerintah terkait.

Populasi dalam penelitian ini mencakup delapan wilayah administratif di Provinsi Banten, terdiri atas empat kabupaten (Pandeglang, Lebak, Tangerang, dan Serang) serta empat kota (Tangerang, Cilegon, Serang, dan Tangerang Selatan). Data yang dikumpulkan mencakup jumlah kasus kriminal yang dilaporkan pada tahun 2023, serta enam variabel independen yang diasumsikan berpengaruh terhadap tingkat kriminalitas, yaitu:

- X_1 : Kepadatan penduduk (jiwa/km²)
- X_2 : Rasio jumlah penduduk terhadap jumlah aparat kepolisian
- X_3 : Tingkat pengangguran terbuka (%)
- X_4 : Angka kemiskinan (%)
- X_5 : Rata-rata lama sekolah (tahun)
- X_6 : Indeks Pembangunan Manusia (IPM) (skala 0–100)

Koordinat geografis (lintang dan bujur) masing-masing wilayah juga dikumpulkan untuk keperluan pemodelan spasial menggunakan Geographically Weighted Regression (GWR).



Gambar 1. Peta Provinsi Banten

Teknik pengumpulan data dilakukan melalui:

1. Studi dokumentasi, yaitu mengunduh dan merekap data dari publikasi resmi BPS dan instansi pemerintah.
2. Ekstraksi data spasial, yaitu pengambilan koordinat geografis pusat wilayah (*centroid*) tiap kabupaten/kota untuk keperluan analisis GWR.
3. Rekapitulasi dan tabulasi data menggunakan Microsoft Excel sebelum dianalisis lebih lanjut.

Analisis data dilakukan melalui beberapa tahap sebagai berikut:

1. Analisis Statistik Deskriptif

Data dianalisis untuk memperoleh gambaran umum distribusi variabel, termasuk nilai minimum, maksimum, rata-rata, simpangan baku, dan varians. Tujuan tahap ini adalah untuk memahami karakteristik awal data dan mendeteksi kemungkinan ketimpangan antarwilayah.

2. Uji Multikolinearitas

Sebelum dilakukan pemodelan regresi, dilakukan uji multikolinearitas terhadap variabel independen dengan menggunakan *Variance Inflation Factor* (VIF). Nilai VIF di atas 10 dianggap menunjukkan adanya korelasi tinggi antarprediktor (Kutner *et al.*, 2005). Variabel dengan VIF tinggi dihapus secara bertahap hingga seluruh prediktor memiliki VIF < 10.

3. Pemodelan Regresi Poisson

Model Regresi Poisson digunakan sebagai model dasar karena sesuai untuk data cacah (*count data*). Estimasi parameter dilakukan dengan metode *Maximum Likelihood Estimation* (MLE), dan kualitas model dinilai melalui deviance serta *Akaike Information Criterion* (AIC) (Cameron & Trivedi, 2013). Rasio deviance terhadap derajat bebas juga dihitung untuk mendeteksi overdispersi.

4. Pemodelan GWR (*Geographically Weighted Regression*)

Model GWR diterapkan untuk menangkap heterogenitas spasial, dengan memungkinkan koefisien regresi berbeda untuk setiap Lokasi (Brunsdon *et al.*, 1996). Estimasi dilakukan dengan

- fungsi pembobot kernel adaptif (*Adaptive Gaussian* dan *Adaptive Bisquare*). Penentuan *bandwidth* optimal dilakukan dengan metode *Cross Validation* (CV) dan minimisasi nilai AIC.
5. Pemodelan Regresi Binomial Negatif (NBR)
Jika overdispersi ditemukan pada model Poisson, maka digunakan Regresi Binomial Negatif (NBR) sebagai alternatif. Model ini menambahkan parameter dispersi untuk mengakomodasi varians yang lebih besar daripada nilai harapan (Hilbe, 2011). Evaluasi model dilakukan berdasarkan AIC, deviance, dan rasio deviance/df.
 6. Perbandingan dan Evaluasi Model
Seluruh model dibandingkan berdasarkan nilai AIC, AICC (AIC dengan koreksi untuk sampel kecil), dan BIC (*Bayesian Information Criterion*). Model dengan nilai informasi terkecil dianggap paling sesuai untuk data yang tersedia. Selain itu, kebermaknaan prediktor dan keterbatasan data juga dibahas secara kritis.

Analisis statistik dilakukan menggunakan perangkat lunak SPSS 22 untuk regresi dasar dan uji multikolinearitas, GWR 4.0 untuk pemodelan spasial GWR, serta Microsoft Excel untuk pengolahan dan visualisasi data awal.

HASIL DAN PEMBAHASAN

Sebelum dilakukan pemodelan, dilakukan analisis deskriptif terhadap seluruh variabel yang digunakan. Hasilnya menunjukkan adanya disparitas yang cukup besar antarwilayah di Provinsi Banten, baik dari segi jumlah kasus kriminalitas maupun variabel sosial ekonomi.

Tabel 1. Analisis Deskriptif

	Min	Max	Rata-Rata	Std. Deviasi	Variansi
Kasus Kriminalitas	421	2648	1177.3750	907.61224	823759.982
Kepadatan Penduduk	438	11012	3806.6250	3884.44752	15088932.55
Rasio jumlah penduduk terhadap jumlah aparat kepolisian	651	3524	1747.7500	900.32768	810589.929
TPT	5.81	9.94	7.5962	1.31308	1.724
Angka Kemiskinan	18200	276330	103266.2500	81176.12368	6589563055
Rata-Rata Lama Sekolah	6.60	11.85	9.1100	1.81405	3.291
IPM	67.68	83.57	75.6713	5.34435	28.562

Berdasarkan Tabel 1, terlihat bahwa jumlah kasus kriminalitas berkisar antara 421 hingga 2.648 kasus, dengan rata-rata 1.177 kasus per wilayah. Kepadatan penduduk bervariasi mulai dari 438 hingga 11.012 jiwa/km², yang menunjukkan perbedaan mencolok antara daerah urban dan rural.

Rasio penduduk terhadap polisi berkisar antara 651 hingga 3.524 orang per petugas, menunjukkan distribusi personel keamanan yang tidak merata. Variabel lainnya seperti tingkat pengangguran, angka kemiskinan, dan rata-rata lama sekolah juga menunjukkan variasi yang cukup tinggi, yang mengindikasikan adanya ketimpangan sosial antarwilayah.



Gambar 2. Grafik Jumlah tindak Kriminalitas di Provinsi Banten Tahun 2023

Uji multikolinearitas dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Hasil awal menunjukkan bahwa beberapa variabel memiliki VIF di atas 10, khususnya rata-rata lama sekolah dan IPM, yang menandakan adanya korelasi tinggi dengan variabel lainnya.

Tabel 2. Nilai VIF awal

No.	Variabel	Nilai VIF
1.	Kepadatan Penduduk	4.804
2.	Rasio jumlah penduduk terhadap jumlah aparat kepolisian	10.273
3.	TPT	3.554
4.	Angka Kemiskinan	10.366
5.	Rata-Rata Lama Sekolah	57.439
6.	IPM	57.517

Setelah dilakukan eliminasi bertahap, diperoleh lima variabel bebas dengan VIF di bawah ambang batas sehingga layak digunakan dalam pemodelan regresi.

Tabel 3. Nilai VIF setelah dilakukan eliminasi

No.	Variabel	Nilai VIF
1.	Kepadatan Penduduk	4.804
2.	Rasio jumlah penduduk terhadap jumlah aparat kepolisian	8.760
3.	TPT	3.535
4.	Angka Kemiskinan	9.669
5.	IPM	6.131

Model regresi Poisson digunakan sebagai pendekatan awal karena sesuai untuk data cacah. Namun, hasil analisis menunjukkan adanya overdispersi, di mana nilai deviance jauh melebihi derajat bebas seperti yang terlihat pada Tabel 4. berikut.

Tabel 4. Nilai Deviansi Model Regresi Poisson

Model	Deviansi
Regresi Poisson dengan GWR	248.624528
Regresi Poisson dengan GLM	4471.764

Tabel 4 menunjukkan bahwa *ratio deviance* terhadap df tercatat sekitar 638, yang mengindikasikan bahwa model ini tidak mampu menangkap variasi data dengan baik. Hal ini sejalan dengan pandangan (Colin & Pravin, 2013). yang menyatakan bahwa model Poisson rentan terhadap overdispersi ketika data menunjukkan variabilitas yang tinggi.

Selain itu, dalam model ini hanya parameter intersep yang signifikan, sedangkan seluruh variabel prediktor lainnya tidak masuk dalam model. Hal ini mengindikasikan lemahnya hubungan global antara variabel-variabel independen dengan tingkat kriminalitas pada data yang tersedia.

Tabel 5. Estimasi Parameter Regresi Poisson

Parameter	Poisson Regression using GWR	Poisson Regression using GLM
β_0	6.056385	7.071
β_1	0.000183	-
β_2	0.000683	-
β_3	-0.023569	-
β_4	-0.000009	-
β_6	-0.000155	-

Berdasarkan Tabel 5, diperoleh nilai estimasi parameter untuk β_0 , β_1 , β_2 , β_3 , β_4 , dan β_6 pada regresi Poisson yang diestimasi menggunakan pendekatan *Geographically Weighted Regression* (GWR). Dengan demikian, model regresinya dapat dituliskan sebagai:

$$\hat{y} = \exp (6.056385 + 0.000183x_1 + 0.000683x_2 - 0.023569x_3 - 0.000009x_4 - 0.000155x_6)$$

Sementara itu, untuk model regresi Poisson menggunakan pendekatan *Generalized Linear Model* (GLM), hanya parameter β_0 (intersep) yang signifikan, sedangkan variabel lainnya (β_1 hingga β_6) tidak signifikan atau dieliminasi dari model sehingga persamaan regresi GLM disederhanakan menjadi:

$$\hat{y} = \exp (7.071)$$

Hasil ini menunjukkan bahwa model GLM tidak mampu menangkap pengaruh prediktor terhadap variabel dependen, karena seluruh prediktor tidak dimasukkan dalam model akibat tidak signifikan secara statistik. Hal ini mengindikasikan adanya masalah dalam kecocokan model (*lack of fit*). Selain itu, hasil ini memperkuat dugaan bahwa data yang digunakan memiliki variasi spasial yang tinggi sehingga pendekatan GLM yang bersifat global kurang sesuai.

Sebaliknya, model GWR memungkinkan setiap lokasi memiliki parameter regresi yang berbeda, sehingga dapat menangkap heterogenitas spasial dalam hubungan antara variabel prediktor dan tingkat kriminalitas. Oleh karena itu, GWR memberikan informasi yang lebih kaya dan relevan secara lokal, serta mampu menunjukkan bagaimana pengaruh setiap variabel prediktor dapat berbeda-beda tergantung pada wilayah geografisnya. Dengan demikian, keberadaan penjelasan ini sangat penting untuk menunjukkan keunggulan model GWR dibandingkan GLM dalam konteks data spasial, serta untuk memberikan alasan metodologis yang kuat atas pemilihan model regresi yang digunakan.

Model GWR diterapkan untuk menangkap variasi spasial dalam hubungan antarvariabel. Model ini menggunakan dua fungsi pembobot kernel adaptif, yaitu Gaussian dan Bisquare. Menariknya, kedua model menghasilkan parameter regresi yang identik di semua wilayah, dan deviasi model tercatat sebagai nol (0.000).

Tabel 6. Nilai Deviansi Setiap Kernel

Kernel	Deviansi
Adaptive bi-square	0.00000000
Adaptive Gaussian	0.00000000

Nilai deviansi sebesar nol mengindikasikan bahwa nilai prediksi model sepenuhnya cocok dengan nilai observasi, yang secara statistik menunjukkan adanya *overfitting*. Dalam konteks pemodelan, deviance nol berarti model sangat sesuai dengan data yang tersedia, tanpa menyisakan variasi tak terjelaskan. Namun, hal ini juga menandakan bahwa model kemungkinan terlalu kompleks untuk jumlah data yang tersedia, atau dalam hal ini, terlalu sensitif terhadap data yang terbatas.

Fenomena ini mengonfirmasi studi Devkota *et al.* (2014) yang menunjukkan bahwa jumlah unit <30 sangat rentan menyebabkan *overfitting* pada GWR, sehingga delapan wilayah seperti dalam studi ini dapat dikatakan berada di bawah ambang optimal.

Tabel 7. Parameter Estimasi setiap wilayah dari kedua kernel

Par	Kernel	Kab. Pandeglang	Kab. Lebak	Kab. Tangerang	Kab. Serang	Tangerang	Cilegon	Serang	Tangerang Selatan
B₀	Gaussian	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385
	Bisquare	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385	6.056385
B₁	Gaussian	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183
	Bisquare	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183	0.000183
B₂	Gaussian	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683
	Bisquare	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683	0.000683
B₃	Gaussian	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357
	Bisquare	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357	-0.02357
B₄	Gaussian	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009
	Bisquare	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009	-0.000009
B₆	Gaussian	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016
	Bisquare	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016	-0.00016

Lebih lanjut, parameter estimasi dari model GWR menggunakan kedua kernel ini menunjukkan hasil yang bukan lagi identik namun sama persis di seluruh wilayah (lihat Tabel 7 dan ringkasan berikut):

$$\hat{y} = \exp (6.056385 + 0.000183x_1 + 0.000683x_2 - 0.02357x_3 - 0.000009x_4 - 0.00016x_6)$$

Model ini berbentuk regresi log-linier, di mana nilai prediksi diperoleh dari fungsi eksponensial dari gabungan linier prediktor. Interpretasi dari setiap koefisien dalam model ini bersifat multiplikatif terhadap ekspektasi jumlah tindak kriminalitas. Secara spesifik, peningkatan satu satuan pada:

- Kepadatan penduduk (x_1) diperkirakan meningkatkan tingkat kriminalitas sebesar 0,0183%;
- Rasio penduduk terhadap polisi (x_2) diperkirakan meningkatkan tingkat kriminalitas sebesar 0,0683%, menunjukkan bahwa semakin banyak penduduk per polisi, maka potensi kriminalitas cenderung meningkat;
- Tingkat Pengangguran Terbuka (x_3) justru menurunkan prediksi kriminalitas sebesar 2,357%, yang tampak kontradiktif namun dapat dijelaskan oleh kemungkinan hubungan tidak langsung, seperti pengangguran yang tidak selalu berujung pada kriminalitas dalam konteks lokal tertentu (Hulu, 2024);
- Angka Kemiskinan (x_4) menurunkan prediksi hanya sebesar 0,0009%, dampak yang sangat kecil dan tidak signifikan;
- IPM (x_6) menurunkan prediksi tingkat kriminalitas sebesar 0,0155%, yang sesuai dengan teori bahwa kualitas hidup yang lebih tinggi dapat mengurangi tekanan sosial dan potensi tindakan kriminal (Handayani, 2017).

Hasil uji signifikansi menunjukkan bahwa variabel x_2 (ratio penduduk terhadap polisi) dan x_3 (TPT) memberikan kontribusi yang signifikan secara statistik terhadap variasi tingkat kriminalitas, di samping intersep. Hal ini menunjukkan bahwa keberadaan aparat keamanan memiliki dampak nyata terhadap penurunan angka kriminalitas, sejalan dengan temuan Hulu (2024) dan Anozi & Novienda (2023) yang menyatakan bahwa distribusi dan ketersediaan sumber daya kepolisian serta dukungan institusional pemerintah mempengaruhi stabilitas sosial di wilayah urban dan semi-urban di Indonesia.

Sementara itu, hubungan negatif yang signifikan antara TPT dan kriminalitas menunjukkan bahwa dalam konteks lokal seperti Provinsi Banten, pengangguran belum tentu meningkatkan kriminalitas. Fenomena ini bisa dijelaskan melalui mekanisme informal seperti pengawasan sosial di komunitas kecil, atau karakteristik pekerjaan informal yang membuat individu tetap terlibat dalam aktivitas ekonomi tanpa tercatat sebagai pekerja formal (Hulu, 2024). Hasil ini sejalan dengan temuan Handayani (2017), yang menunjukkan bahwa variabel kependudukan seperti tingkat pengangguran, kepadatan penduduk, dan partisipasi sekolah memiliki pengaruh signifikan terhadap tingkat kriminalitas di Provinsi Banten. Temuan ini pun mendukung literatur yang menyatakan bahwa pengaruh tingkat pengangguran (TPT) terhadap kriminalitas tidak selalu linier dan dapat bervariasi secara spasial, seperti yang dibahas dalam Devkota *et al.* (2014) dan Páez *et al.* (2011) yang menunjukkan pentingnya analisis lokal dan penggunaan metode spasial untuk menangkap keragaman efek tersebut.

Namun demikian, tidak adanya variasi spasial pada nilai parameter antarwilayah (semuanya identik) menunjukkan bahwa model gagal menangkap heterogenitas lokal, yang seharusnya menjadi keunggulan utama GWR. Temuan ini mengkonfirmasi hasil dari Brunsdon *et al.* (1996) dan Devkota *et al.* (2014), bahwa GWR sangat sensitif terhadap ukuran sampel yang kecil. Dengan hanya delapan unit observasi (kabupaten/kota), GWR tidak mampu menghasilkan koefisien regresi yang berbeda secara lokal. Akibatnya, informasi spasial yang diharapkan dari pendekatan GWR menjadi tidak muncul dalam kasus ini.

Dengan demikian, meskipun GWR secara teoritis menawarkan keunggulan dalam menangkap variasi lokal, hasil penelitian ini menunjukkan bahwa ukuran sampel yang kecil dapat menghambat efektivitasnya, dan bahkan menyebabkan model tampak sempurna secara statistik namun tidak informatif secara substantif.

Tabel 8. Nilai Deviansi Model NBR

Model	Deviansi
NBR	3.878

Sebagai alternatif atas overdispersi, digunakan model Regresi Binomial Negatif (NBR). Hasil estimasi menunjukkan bahwa model ini memiliki rasio deviansi terhadap df sebesar 0,554, yang menunjukkan fit yang sangat baik terhadap data. Nilai AIC juga jauh lebih rendah dibandingkan model Poisson, yakni sebesar 131,14, mengindikasikan bahwa NBR lebih efisien dalam menjelaskan variasi data kriminalitas di Banten. Persamaan regresi yang terbentuk sebagai berikut:

$$\hat{y} = \exp(7.071) \approx 1.175,946$$

Ini menunjukkan bahwa nilai prediksi tetap (konstan) di sekitar 1.176, dengan mengabaikan variabel independen apa pun. Dengan kata lain, model tidak mendeteksi kontribusi signifikan dari prediktor, karena hanya parameter intersep yang signifikan secara statistik.

Fenomena ini serupa dengan yang ditemukan dalam model Poisson, di mana sering kali hanya intersep yang signifikan ketika jumlah sampel terlalu kecil. Dalam kasus ini, ukuran sampel yang terbatas ($n = 8$) kemungkinan menjadi penyebab utama model tidak mampu menangkap pengaruh variabel independen secara memadai. Kondisi ini senada dengan temuan Devkota *et al.* (2014) yang menunjukkan bahwa dataset yang terlalu kecil dapat menyamarkan heterogenitas spasial parameter dalam GWR serta Páez *et al.* (2011) yang memperingatkan bahwa pada sampel kecil GWR rentan menghasilkan korelasi semu dan gagal mengungkap variasi hubungan lokal.

Tabel berikut merangkum perbandingan performa ketiga model yang digunakan:

Tabel 9. Perbandingan Ketiga Model

Model	AIC	Deviance	Rasio Deviance/df
Regresi Poisson	260,62	4471,76	638,82
GWR (Adaptive Gaussian)	0,000	0,000	-
NBR	131,14	3,878	0,554

Berdasarkan tabel 9, dapat disimpulkan bahwa:

- GWR menghasilkan *overfitting*, kemungkinan besar akibat sampel yang terlalu kecil.
- Poisson memberikan fit yang buruk dan mengalami overdispersi berat.
- NBR menunjukkan performa terbaik secara statistik, meskipun tidak menemukan prediktor yang signifikan.

Hasil penelitian ini memperlihatkan bahwa pemodelan spasial seperti GWR tidak selalu dapat diterapkan secara efektif jika jumlah wilayah sangat terbatas. Dalam kondisi seperti ini, model statistik klasik seperti NBR justru memberikan hasil yang lebih stabil, meskipun tidak memperhitungkan variasi spasial. Temuan ini sejalan dengan hasil studi Faradiba, Pradina & Dhuhr (2024) dan Puspitasari (2020) yang menunjukkan bahwa efektivitas GWR sangat bergantung pada cakupan geografis dan ukuran sampel.

Ketidadaan prediktor signifikan dalam model dapat menandakan bahwa kriminalitas di Banten dipengaruhi oleh faktor lain yang belum terukur, atau adanya *noise* akibat keterbatasan data sekunder. Hal ini menunjukkan bahwa selain pendekatan kuantitatif, kajian kualitatif dan pengayaan data juga penting untuk menggali faktor-faktor kontekstual.

KESIMPULAN

Penelitian ini mengevaluasi tiga model regresi, yaitu Regresi Poisson, *Geographically Weighted Regression* (GWR), dan Regresi Binomial Negatif (NBR), dalam menganalisis tingkat kriminalitas di Provinsi Banten dengan menggunakan data dari delapan wilayah administratif. Hasil analisis menunjukkan bahwa model Poisson mengalami overdispersi berat dan gagal menangkap pengaruh variabel prediktor. Model GWR, meskipun secara teori unggul dalam menangkap heterogenitas spasial, menghasilkan parameter yang identik di seluruh wilayah. Hal ini mengindikasikan adanya *overfitting* akibat keterbatasan jumlah unit observasi. Sementara itu, model NBR menunjukkan performa statistik yang paling stabil dan efisien, meskipun tidak menghasilkan prediktor yang signifikan secara statistik.

Temuan ini menegaskan bahwa keberhasilan penerapan metode statistik, khususnya model spasial seperti GWR, sangat bergantung pada karakteristik data, terutama ukuran sampel dan variasi spasial. Dalam kasus data dengan jumlah wilayah terbatas seperti di penelitian ini, model NBR dapat menjadi alternatif yang lebih dapat diandalkan untuk memodelkan data cacah yang overdispersed.

Dari hasil penelitian ini, dapat disimpulkan bahwa:

1. GWR tidak disarankan diterapkan pada jumlah wilayah yang sangat terbatas karena berisiko menghasilkan estimasi koefisien yang seragam dan tidak bermakna secara spasial.
2. Regresi Poisson perlu digunakan dengan hati-hati, terutama ketika data menunjukkan varian yang jauh lebih besar daripada nilai tengahnya.
3. Model NBR layak dipertimbangkan sebagai pendekatan yang *robust* dalam kondisi data yang overdispersed dan bersampel kecil.

REKOMENDASI

Untuk penelitian selanjutnya, disarankan untuk menggunakan jumlah wilayah observasi yang lebih besar dan menyebar secara spasial agar model GWR dapat bekerja optimal. Selain itu, eksplorasi terhadap variabel-variabel lain seperti kriminalitas berdasarkan jenis kasus, atau faktor sosial lain seperti migrasi, tingkat pendidikan informal, dan akses layanan publik juga dapat dipertimbangkan untuk memperkaya analisis. Bagi peneliti, Perlu ada kesadaran bahwa tidak semua metode canggih cocok digunakan untuk semua jenis data. Penggunaan model yang sederhana namun tepat seperti NBR dapat memberikan hasil yang lebih dapat dipercaya dibandingkan model kompleks yang tidak sesuai dengan konteks data. Bagi pengambil kebijakan, hasil ini menyarankan bahwa upaya peningkatan keamanan tidak hanya didasarkan pada sebaran populasi, tetapi juga perlu memperhatikan distribusi aparat dan karakteristik sosial ekonomi setempat.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Fakultas Dakwah Universitas Islam Negeri Sultan Maulana Hasanuddin Banten atas dukungan pendanaan yang telah diberikan untuk pelaksanaan penelitian ini. Dukungan tersebut sangat berperan dalam kelancaran proses pengumpulan data, analisis, hingga penyusunan artikel ilmiah ini.

DAFTAR PUSTAKA

- Anozi, D. T., & Novianda, B. (2023). Economics Development Analysis Journal Socio-Economic and Property Crime Rate in Indonesia. *Economics Development Analysis Journal* Vol, 12(3), 305–318. <http://journal.unnes.ac.id/sju/index.php/edaj>
- Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Excellence in Policing : The 1999 Goldstein Award Winners. *Juvenile Justice*.
- BPS. (2024). STATISTIK KRIMINAL 2024. *Badan Pusat Statistik*.
- Breetzke, G. D. (2018). The concentration of urban crime in space by race: evidence from South Africa. *Urban Geography*, 39(8), 1195–1220. <https://doi.org/10.1080/02723638.2018.1440127>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Colin, C. A., & Pravin, T. (2013). Regression analysis of count data, Second edition. *Regression Analysis of Count Data, Second Edition*, 1–567. <https://doi.org/10.1017/CBO9781139013567>
- Devkota, M., Hatfield, G., & Chintala, R. (2014). Effect of Sample Size on the Performance of Ordinary Least Squares and Geographically Weighted Regression. *British Journal of Mathematics & Computer Science*, 4(1), 1–21. <https://doi.org/10.9734/bjmcs/2014/6050>
- Eck, John and Weisburd, D. L. (2015). Crime Places in Crime Theory. *Crime and Place: Crime Prevention Studies*, 4, 1–33. ssrn: <https://ssrn.com/abstract=2629856>
- Faradiba, Hilda Pradina, Dhuhri, aditya syaiful. (2024). Using the Geographically Weighted Regression (GWR) Method to Find Out the Factors Causing Poverty in North Sumatra Province. *International Journal of Science and Mathematics Education*, 1(1), 18–28.
- Handayani, R. (2017). ANALISIS DAMPAK KEPENDUDUKAN TERHADAP TINGKAT KRIMINALITAS DI PROVINSI BANTEN. *Paper Knowledge . Toward a Media History of Documents*, 12–26.
- Hilbe, J. M. (2011). *Negative Binomial Regression Second Edition*. 1–18. www.cambridge.org
- Hulu, D. M. (2024). Faktor-Faktor Yang Memengaruhi Jumlah Kriminalitas di Indonesia Dengan Regresi Data Panel Pada Tahun 2016-2020. *Indonesian Council of Premier Statistical Science*, 3(2), 37. <https://doi.org/10.24014/icopss.v3i2.32237>
- Páez, A., Farber, S., & Wheeler, D. (2011). A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environment and Planning A*, 43, 2992–3010. <https://doi.org/10.1068/a44111>
- Puspitasari, D. C. (2020). *Pemodelan Persentase Kriminalitas Di Jawa Timur Berdasarkan Pendekatan Geographically Weighted Regression*. [repository.unair.ac.id. https://repository.unair.ac.id/95273/](http://repository.unair.ac.id/95273/)
- Rahman, Y. A., & Prasetyo, A. D. (2018). Economics and Crime Rates in Indonesia. *Jejak*, 11(2), 401–412. <https://doi.org/10.15294/jejak.v11i2.16060>
- Sugiharti, L., Esquivias, M. A., Shaari, M. S., Agustin, L., & Rohmawati, H. (2022). Criminality and Income Inequality in Indonesia. *Social Sciences*, 11(3). <https://doi.org/10.3390/socsci11030142>