

## PEMODELAN JUMLAH SISWA PUTUS SEKOLAH MENGGUNAKAN METODE *GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION*

Eva Selvia, Nurfitri Imro'ah, Wirda Andani

### INTISARI

Jumlah siswa putus sekolah ( $Y$ ) merupakan data cacah sehingga analisis yang tepat untuk memodelkannya adalah dengan regresi Poisson. Namun terdapat asumsi yang harus dipenuhi yaitu equidispersi atau rata-rata harus sama dengan varians. Pada kenyataannya, terdapat suatu kondisi dimana nilai varians lebih besar dari pada nilai rata-rata atau disebut overdispersi. Salah satu alternatif yang dapat digunakan untuk masalah overdispersi pada data cacah adalah regresi Binomial negatif. Regresi Binomial negatif ini kurang tepat jika digunakan pada data yang mengandung heterogenitas spasial atau keragaman antar wilayah. Pengembangan model regresi yang memperhatikan masalah heterogenitas spasial serta masalah overdispersi pada variabel responnya yaitu Geographically Weighted Negative Binomial Regression (GWNBR). Penelitian ini bertujuan untuk memodelkan serta menentukan faktor apa saja yang mempengaruhi jumlah siswa putus sekolah pada jenjang pendidikan dasar menggunakan metode GWNBR. Hasil penelitian yang diperoleh yaitu terdapat dua kelompok kabupaten/kota berdasarkan variabel yang signifikan. Kelompok pertama dipengaruhi oleh semua variabel prediktor, sedangkan kelompok kedua dipengaruhi oleh variabel rasio siswa terhadap guru, tingkat pengangguran terbuka, dan rata-rata pengeluaran perkapita selama sebulan.

**Kata Kunci :** Putus Sekolah, Overdispersi, Heterogenitas Spasial

### PENDAHULUAN

Pendidikan diartikan sebagai usaha manusia dalam menumbuhkan dan mengembangkan potensi-potensi baik jasmani ataupun rohani sesuai dengan nilai-nilai yang berlaku dalam masyarakat dan kebudayaan [1]. Dalam pembukaan UUD 1945 tercantum tujuan Indonesia yakni mencerdaskan kehidupan bangsa sehingga Indonesia mengupayakan berbagai peraturan dan kebijakan untuk memenuhi hak setiap warga negara dalam mendapatkan layanan pendidikan. Namun masih terdapat masalah dalam bidang pendidikan yaitu sejumlah siswa yang tidak dapat melanjutkan pendidikan atau putus sekolah.

Putus sekolah adalah sebutan untuk mantan peserta didik yang tidak dapat menyelesaikan suatu jenjang pendidikan, sehingga tidak dapat melanjutkan studi ke jenjang pendidikan yang lebih tinggi. Berdasarkan data dari Dinas Pendidikan dan Kebudayaan Provinsi Kalimantan Barat, pada tahun ajaran 2020/2021 terdapat sebanyak 1.011 siswa yang putus sekolah pada jenjang pendidikan dasar. Putus sekolah digunakan sebagai salah satu indikator yang mengukur kemajuan sumber daya manusia dalam bidang pendidikan. Untuk mengurangi jumlah siswa yang putus sekolah perlu diketahui faktor apa saja yang menyebabkan siswa putus sekolah.

Metode statistika yang dapat digunakan untuk menguji hubungan antara variabel respon dengan beberapa variabel prediktor yaitu analisis regresi. Apabila variabel respon ( $Y$ ) yang digunakan berupa data cacah, maka dapat menggunakan regresi Poisson dimana terdapat syarat yang harus terpenuhi yakni nilai rata-rata dan varians harus sama. Salah satu pelanggaran dalam regresi Poisson yaitu overdispersi atau kondisi dimana nilai variansnya lebih besar dari pada rata-ratanya. Metode yang dapat digunakan sebagai alternatif jika terdapat masalah overdispersi pada regresi Poisson adalah regresi Binomial negatif [2].

Setiap daerah mempunyai kondisi geografis, sosial budaya dan ekonomi berbeda-beda. Hal tersebut menandakan bahwa ada efek heterogenitas spasial antar wilayah. Sehingga diperlukan suatu model regresi yang memperhitungkan faktor lokasi (spasial) [3]. Pengembangan model regresi Binomial negatif yang memperhatikan faktor heterogenitas spasial adalah *Geographically Weighted Negative Binomial Regression* (GWNBR) [4]. Tujuan penelitian ini yaitu menentukan faktor apa saja yang mempengaruhi siswa putus sekolah pada jenjang pendidikan dasar setiap kabupaten/kota serta memodelkan jumlah siswa putus sekolah menggunakan metode *Geographically Weighted Negative Binomial Regression* (GWNBR) dengan batasan masalah yaitu jumlah siswa putus sekolah pada jenjang pendidikan dasar dan fungsi pembobot yang digunakan yaitu fungsi *kernel adaptive bivariate*.

Langkah pertama dalam penelitian ini yaitu mendeskripsikan data jumlah siswa putus sekolah menggunakan statistik deskriptif, mengecek unsur multikolinearitas antar variabel prediktor, memodelkan data penelitian yang digunakan menggunakan regresi Poisson kemudian melakukan pendeteksian apakah model regresi Poisson mengalami overdispersi. Apabila didalam model regresi Poisson terdapat overdispersi, maka di atasi dengan menggunakan regresi Binomial negatif. Selanjutnya melakukan uji heterogenitas spasial untuk mengetahui apakah terdapat keragaman spasial antar lokasi pengamatan. Jika terdapat unsur heterogenitas spasial, maka dilakukan pemodelan menggunakan metode *Geographically Weighted Negative Binomial Regression*.

### Analisis Regresi

Analisis regresi adalah metode yang digunakan untuk memperkirakan nilai suatu variabel berdasarkan nilai variabel yang lain. Secara umum, model analisis regresi dinyatakan [5]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

dengan  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  merupakan nilai yang belum diketahui (parameter) dan harus diduga, sedangkan  $\varepsilon$  merupakan lambang sisaan.

Analisis regresi terdiri dari regresi linear dan regresi nonlinear. Analisis regresi linear mempunyai asumsi kelinearan dalam parameter dan berdistribusi normal. Jika model regresi dari data yang digunakan tidak linear secara parameter dan berdistribusi normal maka analisis regresi yang digunakan adalah analisis regresi nonlinear. Untuk memodelkan data yang tidak linear secara parameter menggunakan *Generalized Linear Model* (GLM) [5].

### Multikolinearitas

Multikolinearitas adalah suatu kondisi di dalam model regresi dimana variabel prediktor satu mempunyai hubungan atau korelasi dengan variabel prediktor lainnya. Pengecekan multikolinearitas dapat dilakukan melalui nilai *Variance Inflation Factors* (VIF) dengan kriteria penilaian yaitu jika nilai VIF lebih besar dari 10 maka dapat disimpulkan bahwa model tersebut mengandung unsur multikolinearitas. Untuk mendapatkan nilai VIF dapat dihitung dengan rumus berikut [6]:

$$VIF_k = \frac{1}{1-R_k^2} \quad (1)$$

dengan  $R_k^2$  adalah nilai koefisien determinasi variabel  $X_k$ .

### Regresi Poisson

Salah satu model yang sering digunakan untuk memodelkan variabel respon berupa data cacah atau jumlahan adalah Regresi Poisson. Jika variabel random diskrit ( $Y$ ) diasumsikan berdistribusi Poisson dengan parameter  $\mu$  maka fungsi peluangnya dinyatakan [6]:

$$P(y|\mu) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots, n$$

dengan  $\mu$  merupakan rata-rata variabel respon yang berdistribusi Poisson. Model regresi Poisson dinyatakan [7]:

$$\ln(\mu_i) = \beta_0 + \sum_{k=1}^p \beta_k x_{ik}, i = 1, 2, \dots, n \quad (2)$$

Karena variabel respon diasumsikan berdistribusi Poisson maka model menjadi [7]:

$$\mu_i = \exp(x_i^T \beta) \tag{3}$$

dengan  $x_i^T = [x_{1i}, x_{2i}, \dots, x_{ki}]$  dan  $\beta = [\beta_1, \beta_2, \dots, \beta_k]$

Proses estimasi parameter regresi Poisson menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan fungsi *likelihood*nya yaitu:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(-\mu_i)(\mu_i)^{y_i}}{y_i!}$$

Kemudian fungsi *likelihood* tersebut diubah menjadi *log-likelihood* Poisson menjadi:

$$\ln L(\beta) = - \sum_{i=1}^n \exp(x_i^T \beta) + \sum_{i=1}^n y_i x_i^T \beta - \sum_{i=1}^n \ln(y_i!)$$

Persamaan *log-likelihood* tersebut diturunkan terhadap  $\beta$  serta disamakan dengan nol. Namun persamaan tersebut masih berbentuk implisit sehingga digunakan alternatif lain untuk menentukan nilai prosedur MLE yaitu menggunakan metode iterasi *Fisher-Scoring* [8].

Proses uji signifikansi parameter regresi Poisson terbagi menjadi dua yaitu pengujian parameter secara serentak dan parsial. Pengujian signifikansi secara serentak dilakukan menggunakan *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis pengujian sebagai berikut [9]:

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

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

Statistik uji yang digunakan yaitu:

$$D(\hat{\beta}) = -2 \ln \left( \frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 \left( \ln L(\hat{\Omega}) - \ln L(\hat{\omega}) \right) \tag{4}$$

Kriteria pengujian signifikansi parameter model regresi Poisson secara serentak yaitu tolak  $H_0$  jika  $D(\hat{\beta}) > \chi^2_{(k;\alpha)}$ .

Proses selanjutnya yaitu uji signifikansi secara parsial. Hipotesis pengujian signifikansi parameter secara parsial adalah [9]:

$$H_0: \beta_k = 0 \quad (\text{pengaruh variabel } k \text{ tidak signifikan})$$

$$H_1: \beta_k \neq 0; k = 1, 2, \dots, p \quad (\text{pengaruh variabel } k \text{ signifikan})$$

Statistik uji yang digunakan yaitu:

$$Z_{hitung} = \frac{\hat{\beta}_k}{SE\hat{\beta}_k} \tag{5}$$

dengan  $SE\hat{\beta}_k$  adalah nilai standar error dari parameter  $\beta_k$  serta kriteria pengujiannya adalah tolak  $H_0$  apabila  $|Z_{hitung}| > Z_{\alpha/2}$ .

**Overdispersi**

Overdispersi adalah suatu kondisi dimana nilai varians lebih besar dari nilai rata-ratanya. Hal ini melanggar asumsi regresi Poisson yang mengharuskan nilai varians sama dengan nilai rata-rata. Apabila didalam model regresi Poisson mengalami overdispersi namun tetap menggunakan regresi Poisson maka kesimpulan yang diperoleh menjadi tidak valid dikarenakan nilai standar errornya menjadi *underestimate*. Untuk mendeteksi adanya overdispersi dalam model regresi Poisson dapat dilihat dari nilai *devians* yang dibagi derajat bebasnya. Apabila nilai tersebut lebih besar dari satu maka disimpulkan bahwa terjadi overdispersi pada data. Jika terjadi kasus overdispersi, maka metode alternatif yang dapat digunakan yaitu regresi Binomial negatif [10].

**Regresi Binomial Negatif**

Regresi Binomial negatif merupakan alternatif untuk menangani adanya kasus overdispersi pada regresi Poisson yang merupakan regresi campuran Poisson-Gamma [10].

Misalkan:

$Y \sim \text{Poisson}(\mu)$  dengan fungsi peluang sebagai berikut:  $\frac{e^{-\mu} \mu^y}{y!}$

$\mu \sim \text{Gamma}(\alpha, \beta)$  dengan fungsi peluang sebagai berikut:  $\frac{1}{\Gamma(\alpha)\beta^\alpha} \mu^{\alpha-1} \exp\left(-\frac{\mu}{\beta}\right)$

Maka fungsi massa peluang Binomial negatif dinyatakan [8]:

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

dengan  $y = 0, 1, 2, \dots, n; \mu = \exp(x_i^T \beta)$

Model regresi Binomial negatif yaitu:

$$\mu_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ik}) \quad (6)$$

dengan  $\mu_i$  merupakan rata-rata jumlah kejadian yang terjadi dalam jangka waktu tertentu. Estimasi parameter regresi Binomial negatif menggunakan metode *Maximum Likelihood Estimation* (MLE). Fungsi *likelihood* dari regresi Binomial negatif sebagai berikut [2]:

$$L(\beta, \theta) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \frac{1}{\theta})}{y_i! \Gamma(\frac{1}{\theta})} \left(\frac{1}{1 + \theta \exp(x_i^T \beta)}\right)^{\frac{1}{\theta}} \left(\frac{\theta \exp(x_i^T \beta)}{1 + \theta \exp(x_i^T \beta)}\right)^{y_i} \right\}$$

Selanjutnya yaitu membentuk fungsi log dari fungsi *likelihood* sehingga menjadi:

$$\ln L(\beta, \theta) = \sum_{i=1}^n \left\{ \ln \frac{\Gamma(y_i + \frac{1}{\theta})}{y_i! \Gamma(\frac{1}{\theta})} + \theta^{-1} \ln \left(\frac{1}{1 + \theta \exp(x_i^T \beta)}\right) + y_i \ln \left(\frac{\theta \exp(x_i^T \beta)}{1 + \theta \exp(x_i^T \beta)}\right) \right\}$$

Kemudian fungsi *log-likelihood* diturunkan terhadap masing-masing parameter  $\beta$  dan  $\theta$  kemudian disamakan dengan nol dan diselesaikan dengan iterasi *Fisher-Scoring*.

Proses selanjutnya yaitu pengujian signifikansi parameter yang terdiri dari uji signifikansi secara serentak dan secara parsial. Pengujian signifikansi parameter model regresi Binomial negatif secara serentak maupun parsial memiliki metode atau cara yang sama dengan regresi Poisson.

### Aspek Spasial

Untuk melihat apakah data pada penelitian layak untuk dianalisis menggunakan regresi spasial atau tidak perlu dilakukan pengujian aspek spasial. Pengujian aspek spasial dilakukan melalui uji heterogenitas spasial [8]. Heterogenitas spasial merupakan variasi pada tiap lokasi. Pengujian heterogenitas spasial dapat dilakukan melalui uji *Breusch-Pagan* (BP) dengan hipotesis sebagai berikut [4]:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \quad (\text{variansi antarlokasi sama})$$

$$H_1: \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2, \text{ dimana } i = 1, 2, \dots, n \quad (\text{variansi antarlokasi berbeda})$$

Statistik uji yang digunakan untuk uji *Breusch-Pagan* adalah [4]:

$$BP = \left(\frac{1}{2}\right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \quad (7)$$

Keterangan:

$\mathbf{f}$  :  $(f_1, f_2, \dots, f_n)^T$  dengan  $f_i = \frac{e_i^2}{\sigma^2} - 1$

$\mathbf{Z}$  : matriks ukuran  $n \times (k + 1)$  berisi vektor yang sudah dinormalkan ( $\mathbf{Z}$ ) untuk setiap pengamatan, dengan  $k$  adalah banyaknya variabel predictor

$e_i$  :  $(y_i - \hat{y}_i)$ , merupakan residual pengamatan ke- $i$

$\sigma^2$  : varians sampel  $y$

Kriteria pengujiaanya yaitu tolak  $H_0$  jika nilai  $BP > \chi_{(k;\alpha)}^2$  atau P-value  $< \alpha$  yang berarti terjadi heterogenitas spasial atau variansi antarlokasi berbeda.

**Fungsi Pembobot Optimum**

Pembobotan untuk setiap area diperlukan karena setiap lokasi atau area mempunyai variasi atau keragaman yang berbeda-beda. Pada penelitian ini pembobotan dilakukan menggunakan fungsi *kernel adaptive bisquare* yang mempunyai *bandwidth* berbeda untuk setiap area yang berbeda pula. *Bandwidth* berfungsi sebagai dasar penentuan bobot setiap area pengamatan terhadap model regresi pada area tersebut. Untuk mendapatkan pembobotan dengan fungsi *kernel adaptive bisquare* menggunakan rumus berikut [10]:

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

Keterangan:

- $i, j$  : lokasi pengamatan ke-1,2, ...,  $n$  dimana  $i \neq j$
- $c_i$  : nilai *bandwidth* optimum untuk lokasi pengamatan ke- $i$
- $d_{ij}$  : jarak *Euclidean* antar lokasi pengamatan ke- $i$  dan lokasi pengamatan ke- $j$

Untuk mendapatkan nilai  $d_{ij}$  menggunakan rumus berikut:

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

Keterangan:

- $u_i$  : koordinat lintang (latitude) lokasi pengamatan ke- $i$
- $u_j$  : koordinat lintang (latitude) lokasi pengamatan ke- $j$
- $v_i$  : koordinat bujur (longitude) lokasi pengamatan ke- $i$
- $v_j$  : koordinat bujur (longitude) lokasi pengamatan ke- $j$

Perhitungan nilai *bandwidth* optimum menggunakan kriteria *cross validation* (CV) dengan rumus sebagai berikut:

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

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

***Geographically Weighted Negative Binomial Regression***

*Geographically Weighted Negative Binomial Regression* (GWNBR) adalah salah satu metode yang dapat digunakan untuk memprediksi data cacah yang mempunyai masalah overdispersi serta heterogenitas spasial. GWNBR menghasilkan parameter lokal dimana setiap lokasi memiliki parameter yang berbeda-beda atau dapat dikatakan GWNBR merupakan pengembangan dari regresi Binomial negatif. Model GWNBR dinyatakan [11]:

$$\mu_i = \exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{1i} + \dots + \beta_k(u_i, v_i)X_{ik}) \tag{8}$$

Estimasi parameter model GWNBR menggunakan metode *Maximum Likelihood Estimation* (MLE). Persamaan *likelihood* untuk GWNBR adalah sebagai berikut [12]:

$$L(.) = \prod_i^n \frac{\Gamma(y_i + \theta(u_i, v_i)^{-1})}{y_i! (\Gamma(\theta(u_i, v_i)^{-1}))} \left( \frac{1}{1 + \theta(u_i, v_i) \exp(x_i^T \beta(u_i, v_i))} \right)^{\frac{1}{\theta}} \left( \frac{\theta(u_i, v_i) \exp(x_i^T \beta(u_i, v_i))}{1 + \theta(u_i, v_i) \exp(x_i^T \beta(u_i, v_i))} \right)^{y_i}$$

dengan  $L(.) = L(\beta(u_i, v_i), \theta_i | y_i, x_i)$ .

Fungsi *likelihood* kemudian diubah ke dalam bentuk *log-likelihood* yang diberi pembobot letak geografis menjadi seperti berikut [3]:

$$\ln L(.) = \sum_{i=1}^n w_{ij}(u_i, v_i) \left\{ \ln \frac{\Gamma(y_i + \theta^{-1})}{y! \Gamma(\theta^{-1})} - (y_i + \theta^{-1}) \ln(1 + \theta \mu_i) + y_i \ln(\theta \mu_i) \right\}$$

Kemudian fungsi *log-likelihood* diturunkan terhadap masing-masing parameter yaitu  $\beta$  dan  $\theta$ , disamakan dengan nol dan diselesaikan dengan iterasi *Fisher-Scoring*. Metode tersebut digunakan untuk memperoleh solusi dari fungsi *log-likelihood* sehingga didapatkan nilai yang cukup konvergen sebagai estimasi bagi tiap parameter [11].

Proses selanjutnya yaitu melakukan pengujian signifikansi parameter model GWNBR yang terdiri dari pengujian signifikansi parameter secara serentak dan parsial. Hipotesis uji untuk pengujian signifikansi parameter secara serentak yaitu [4]:

$$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \beta_k(u_i, v_i) = 0 \text{ dimana } i = 1, 2, \dots, n.$$

$$H_1: \text{minimal ada satu } \beta_k(u_i, v_i) \neq 0, \text{ dimana } k = 1, 2, \dots, p \text{ dan } i = 1, 2, \dots, n.$$

Statistik uji yang digunakan untuk pengujian signifikansi secara serentak untuk model GWNBR adalah sebagai berikut:

$$D(\hat{\beta}(u_i, v_i)) = -2 \ln \left( \frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 \left( \ln L(\hat{\Omega}) - \ln L(\hat{\omega}) \right) \quad (9)$$

Kriteria pengujianya yaitu dengan menggunakan taraf signifikansi  $\alpha$ , maka tolak  $H_0$  jika  $D(\hat{\beta}(u_i, v_i)) > \chi_{(k; \alpha)}^2$  atau jika *P-value*  $< \alpha$ .

Apabila pada uji signifikansi parameter secara serentak diperoleh keputusan tolak  $H_0$ , maka dilanjutkan dengan uji signifikansi parameter secara parsial dengan hipotesis sebagai berikut [4]:

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

$$H_1: \beta_k(u_i, v_i) \neq 0, \text{ dimana } k = 1, 2, \dots, p \text{ dan } i = 1, 2, \dots, n.$$

Statistik uji untuk pengujian signifikansi parameter secara parsial adalah sebagai berikut:

$$Z_{hitung} = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))} \quad (10)$$

Kriteria pengujian signifikansi secara parsial dengan menggunakan taraf signifikansi  $\alpha$  yaitu tolak  $H_0$  jika  $|Z_{hitung}| > Z_{\alpha/2}$  atau jika *P-value*  $< \alpha$  [8].

## HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Pendidikan dan Kebudayaan Provinsi Kalimantan Barat dan publikasi dari Badan Pusat Statistik. Data yang digunakan terdiri dari variabel respon ( $Y$ ) yaitu jumlah anak putus sekolah pada jenjang pendidikan dasar (SD), sedangkan variabel prediktor terdiri dari rasio siswa terhadap guru ( $X_1$ ), jumlah penduduk miskin ( $X_2$ ), tingkat pengangguran terbuka ( $X_3$ ), dan rata-rata pengeluaran perkapita selama sebulan ( $X_4$ ). Penelitian ini berfokus pada setiap kabupaten/kota di Provinsi Kalimantan Barat pada tahun 2021 dengan jumlah data sebanyak 14 serta menggunakan bantuan *software* R studio.

Statistik deskriptif dari data jumlah siswa putus sekolah pada jenjang pendidikan dasar beserta faktor-faktor yang diduga mempengaruhinya adalah sebagai berikut:

**Tabel 1.** Statistik Deskriptif

	Rata-Rata	Maksimum	Minimum	Varians
$Y$	72,21	290	14	5322,95
$X_1$	15,04	20,27	10,72	8,76
$X_2$	26.304	53.040	10.720	177.788.797,8
$X_3$	5,38	12,38	2,66	7,83
$X_4$	1.143.016	1.694.619	941.823	48.163.510.546

Tabel 1 menunjukkan bahwa rata-rata jumlah siswa yang putus sekolah di Provinsi Kalimantan Barat adalah sebanyak 72,21 atau 72 orang dimana jumlah siswa putus sekolah tertinggi yaitu Kota Pontianak sebanyak 290 dan terendah yaitu Kabupaten Melawi sebanyak 14.

**Uji Multikolinearitas**

Pengecekan multikolinearitas dilakukan guna melihat hubungan atau korelasi antar variabel prediktor. Apabila nilai VIF yang diperoleh lebih besar dari 10, maka dapat disimpulkan bahwa terdapat unsur multikolinearitas antar variabel prediktor. Berikut hasil dari pengecekan multikolinearitas antar variabel prediktor yang diduga mempengaruhi jumlah siswa putus sekolah pada jenjang pendidikan dasar:

**Tabel 2.** Nilai VIF masing-masing variabel prediktor

Variabel	$X_1$	$X_2$	$X_3$	$X_4$
VIF	1,446	2,606	4,576	2,652

Berdasarkan Tabel 2 diperoleh hasil bahwa semua nilai VIF lebih kecil dari 10 sehingga dapat dikatakan bahwa tidak terdapat unsur multikolinearitas antar variabel prediktor yang diduga mempengaruhi jumlah anak putus sekolah pada jenjang pendidikan dasar di Provinsi Kalimantan Barat.

**Pemodelan Regresi Poisson**

Hasil estimasi parameter dengan menggunakan model regresi Poisson untuk data jumlah siswa putus sekolah pada jenjang pendidikan dasar disajikan dalam Tabel 3 berikut:

**Tabel 3.** Estimasi Parameter Model Regresi Poisson

Variabel	Estimasi	Standar Error	$ Z_{hitung} $	<i>P-Value</i>
<i>Intercept</i>	1,924	0,249	7,712	$1,24 \times 10^{-14}$
$X_1$	$-3,576 \times 10^{-2}$	$1,486 \times 10^{-2}$	-2,407	0,016
$X_2$	$2,645 \times 10^{-5}$	$2,395 \times 10^{-6}$	11,046	$2 \times 10^{-16}$
$X_3$	0,128	$1,970 \times 10^{-2}$	6,518	$7,12 \times 10^{-11}$
$X_4$	$1,139 \times 10^{-6}$	$2,474 \times 10^{-7}$	4,602	$4,18 \times 10^{-6}$
<i>Devians: 232,11</i>				DF: 9

Tabel 3 menunjukkan nilai *devians* sebesar 232,11 yang lebih besar dari nilai  $\chi^2_{(4;0,1)} = 7,779$  maka dapat disimpulkan tolak  $H_0$  yang artinya minimal terdapat satu parameter yang mempengaruhi variabel jumlah siswa putus sekolah pada jenjang pendidikan dasar. Oleh karena itu, pengujian signifikansi secara parsial dapat dilanjutkan.

Pengujian signifikansi secara parsial dengan taraf signifikansi  $\alpha$  sebesar 10% diperoleh nilai  $Z_{(0,1/2)} = 1,64$ . Nilai tersebut dibandingkan dengan nilai  $|Z_{hitung}|$  masing-masing variabel prediktor. Berdasarkan hasil pengujian diperoleh hasil bahwa semua variabel prediktor memiliki  $|Z_{hitung}|$  lebih besar dari 1,64 yang artinya semua variabel berpengaruh signifikan terhadap jumlah siswa putus sekolah pada jenjang pendidikan dasar.

Berdasarkan Tabel 3 diketahui bahwa nilai *devians* regresi Poisson sebesar 232,11 dengan derajat bebasnya yaitu 9 sehingga rasio nilai *devians* dengan derajat bebasnya yaitu 25,79. Nilai tersebut lebih besar dari 1 maka dapat disimpulkan bahwa data jumlah siswa putus sekolah pada jenjang pendidikan dasar mengalami overdispersi. Maka dari itu, regresi Poisson tidak sesuai untuk memodelkan data tersebut sehingga untuk mengatasi masalah tersebut maka akan digunakan regresi Binomial negatif.

**Pemodelan Regresi Binomial Negatif**

Hasil estimasi parameter regresi Binomial negatif dapat dilihat pada Tabel 4 berikut:

**Tabel 4.** Estimasi Parameter Model Regresi Binomial Negatif

Variabel	Estimasi	Standar Error	$ Z_{hitung} $	P-Value
<i>Intercept</i>	2,036	1,435	1,419	0,189
$X_1$	$-3,832 \times 10^{-2}$	0,084	0,455	0,659
$X_2$	$2,844 \times 10^{-5}$	$1,521 \times 10^{-5}$	1,870	0,094
$X_3$	0,102	0,106	0,967	0,358
$X_4$	$1,163 \times 10^{-6}$	$1,323 \times 10^{-6}$	0,879	0,402
<i>Devians: 14,537</i>				DF: 9

Berdasarkan Tabel 4 diketahui bahwa nilai *devians* regresi Binomial negatif yaitu sebesar 14,537 yang lebih besar dari  $\chi^2_{(4;0,1)} = 7,779$  sehingga dapat disimpulkan tolak  $H_0$  yang artinya minimal terdapat satu parameter yang mempengaruhi variabel jumlah siswa putus sekolah pada jenjang pendidikan dasar. Maka dari itu, dapat dilanjutkan dengan pengujian signifikansi secara parsial.

Berdasarkan hasil pengujian secara parsial dengan taraf signifikansi 10% diperoleh nilai  $Z_{(0,1/2)} = 1,64$ . Nilai tersebut dibandingkan dengan nilai  $|Z_{hitung}|$  masing-masing variabel prediktor. Berdasarkan hasil pengujian diperoleh hasil bahwa variabel prediktor yang memiliki  $|Z_{hitung}|$  lebih besar dari 1,64 adalah variabel jumlah penduduk miskin ( $X_2$ ).

Rasio nilai *devians* regresi Binomial negatif dengan derajat bebasnya yaitu 1,615. Nilai tersebut mendekati 1 dan lebih kecil dari pada rasio nilai *devians* model regresi Poisson yaitu 25,79. Maka dari itu, dapat dikatakan bahwa regresi Binomial negatif lebih baik digunakan untuk memodelkan data diskrit yang mengandung overdispersi dari pada regresi Poisson.

### Uji Heterogenitas Spasial

Berdasarkan hasil pengujian heterogenitas spasial, diperoleh nilai BP sebesar 13,48 yang lebih besar dari  $\chi^2_{(4;0,1)} = 7,779$  dan nilai *P-value* sebesar 0,009 yang lebih kecil dari nilai  $\alpha = 0,1$  artinya tolak  $H_0$  atau dapat disimpulkan bahwa terdapat perbedaan karakteristik antar lokasi pengamatan satu dengan yang lainnya. Karena terdapat heterogenitas spasial antar wilayah di Provinsi Kalimantan Barat maka metode *Geographically Weighted Negative Binomial Regression* dapat digunakan dalam analisis selanjutnya.

### Pemodelan *Geographically Weighted Negative Binomial Regression*

Berdasarkan pengujian signifikansi secara serentak diperoleh nilai *devians* model GWNBR sebesar 322,845 yang lebih besar dari  $\chi^2_{(4;0,1)} = 7,779$  sehingga disimpulkan bahwa  $H_0$  ditolak yang artinya terdapat paling sedikit ada satu parameter model GWNBR yang berpengaruh signifikan terhadap variabel jumlah siswa putus sekolah pada jenjang pendidikan dasar. Maka dari itu perlu dilanjutkan dengan pengujian secara parsial untuk mengetahui faktor apa saja yang berpengaruh terhadap variabel respon pada masing-masing kabupaten/kota.

Dengan menggunakan taraf signifikansi sebesar 10% diperoleh nilai  $Z_{(0,1/2)} = 1,64$ . Jika nilai  $|Z_{hitung}| > 1,64$  maka tolak  $H_0$  yang artinya variabel tersebut berpengaruh signifikan terhadap model. Berdasarkan variabel yang signifikan dapat dibentuk dua kelompok kabupaten/kota sebagai berikut:

**Tabel 5.** Kelompok Kabupaten/Kota Berdasarkan Variabel yang Signifikan

No	Kabupaten/Kota	Variabel yang signifikan
1	Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Kubu Raya, Pontianak, dan Singkawang	$X_1, X_2, X_3, X_4$
2	Sintang, Kapuas Hulu, Sekadau, Melawi, dan Kayong Utara	$X_1, X_3, X_4$

Berdasarkan Tabel 5 diketahui bahwa di Kalimantan Barat terdapat dua kelompok Kabupaten Kota berdasarkan variabel yang mempengaruhinya. Kelompok pertama dipengaruhi oleh semua variabel



prediktor, sedangkan kelompok yang kedua dipengaruhi oleh variabel rasio siswa terhadap guru, tingkat pengangguran terbuka dan rata-rata pengeluaran perkapita selama sebulan. Pada kelompok kedua, jumlah penduduk miskin mengalami kenaikan dibanding tahun sebelumnya namun hal tersebut tidak mempengaruhi banyaknya siswa yang putus sekolah pada jenjang pendidikan dasar pada lokasi tersebut.

Berikut disajikan koefisien parameter masing-masing kabupaten/kota di Provinsi Kalimantan Barat.

**Tabel 6.** Koefisien Parameter Setiap Kabupaten/Kota

<i>i</i>	Kab/Kota	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
1	Sambas	$4,54 \times 10^{-11}$	$2,81 \times 10^{-10}$	$7,70 \times 10^{-6}$	$-2,13 \times 10^{-10}$	$8,12 \times 10^{-7}$
2	Bengkayang	$1,47 \times 10^{-11}$	$6,79 \times 10^{-11}$	$4,55 \times 10^{-6}$	$6,65 \times 10^{-11}$	$9,16 \times 10^{-7}$
3	Landak	$3,01 \times 10^{-11}$	$4,50 \times 10^{-10}$	$6,28 \times 10^{-6}$	$-1,55 \times 10^{-10}$	$8,94 \times 10^{-7}$
4	Mempawah	$1,42 \times 10^{-10}$	$1,16 \times 10^{-9}$	$1,33 \times 10^{-5}$	$-1,38 \times 10^{-9}$	$6,71 \times 10^{-7}$
5	Sanggau	$1,39 \times 10^{-10}$	$1,69 \times 10^{-9}$	$1,29 \times 10^{-5}$	$-8,55 \times 10^{-10}$	$7,12 \times 10^{-7}$
6	Ketapang	$3,19 \times 10^{-11}$	$-1,72 \times 10^{-10}$	$6,36 \times 10^{-6}$	$2,16 \times 10^{-11}$	$8,07 \times 10^{-7}$
7	Sintang	$2,29 \times 10^{-12}$	$7,35 \times 10^{-11}$	$1,58 \times 10^{-6}$	$1,04 \times 10^{-10}$	$1,07 \times 10^{-6}$
8	Kapuas Hulu	$8,34 \times 10^{-12}$	$2,78 \times 10^{-10}$	$3,21 \times 10^{-6}$	$1,02 \times 10^{-10}$	$1,04 \times 10^{-6}$
9	Sekadau	$1,78 \times 10^{-11}$	$1,83 \times 10^{-10}$	$4,85 \times 10^{-6}$	$1,56 \times 10^{-10}$	$9,43 \times 10^{-7}$
10	Melawi	$4,93 \times 10^{-12}$	$5,05 \times 10^{-11}$	$2,61 \times 10^{-6}$	$1,50 \times 10^{-10}$	$1,02 \times 10^{-6}$
11	Kayong Utara	$1,47 \times 10^{-12}$	$8,79 \times 10^{-13}$	$-3,81 \times 10^{-7}$	$-2,50 \times 10^{-11}$	$1,12 \times 10^{-6}$
12	Kubu Raya	$2,45 \times 10^{-10}$	$2,05 \times 10^{-9}$	$1,72 \times 10^{-5}$	$-2,64 \times 10^{-9}$	$5,87 \times 10^{-7}$
13	Kota Pontianak	$2,30 \times 10^{-10}$	$1,90 \times 10^{-9}$	$1,67 \times 10^{-5}$	$-2,44 \times 10^{-9}$	$5,90 \times 10^{-7}$
14	Kota Singkawang	$1,06 \times 10^{-10}$	$7,79 \times 10^{-10}$	$1,15 \times 10^{-5}$	$-8,83 \times 10^{-10}$	$7,03 \times 10^{-7}$

Berdasarkan tabel 6, berikut disajikan contoh model GWNBR dengan menggunakan persamaan (8) pada lokasi pengamatan ke-8 atau Kabupaten Kapuas Hulu:

$$\mu_8 = \exp(8,341 \times 10^{-11} + 2,78 \times 10^{-10}X_1 + 1,02 \times 10^{-10}X_3 + 1,04 \times 10^{-6}X_4)$$

**KESIMPULAN**

Berdasarkan hasil analisis dan pembahasan pada penelitian ini dapat disimpulkan bahwa regresi Binomial negatif lebih baik digunakan dari pada regresi Poisson karena menghasilkan nilai rasio *devians* yang lebih kecil. Namun terdapat efek heterogenitas spasial sehingga dilakukan pemodelan menggunakan metode *Geographically Weighted Negative Binomial Regression* dengan fungsi pembobot yang digunakan yaitu *kernel adaptive bivariate*. Hasil yang diperoleh yaitu terdapat dua kelompok kabupaten/kota berdasarkan variabel yang berpengaruh signifikan terhadap jumlah siswa putus sekolah pada jenjang pendidikan dasar.

**DAFTAR PUSTAKA**

[1] A. Rahman, S. A. Munandar, A. Fitriani, Y. Karlina and Yumriani, "Pengertian Pendidikan, Ilmu Pendidikan dan Unsur-Unsur Pendidikan," *Al Urwatul Wutsqa: Kajian Pendidikan Islam*, vol. 2 (1), pp. 1-8, 2022.

[2] A. Sauddin, N. I. Auliah and W. Alwi, "Pemodelan Jumlah Kematian Ibu di Provinsi Sulawesi Selatan Menggunakan Regresi Binomial Negatif," *Jurnal Matematika dan Statistika Serta Aplikasinya*, vol. 8(2), pp. 42-47, 2020.

- [3] L. E. Afri, A. and A. Djuraidah, "Model Regresi Binomial Negatif Terboboti Geografis untuk Data Kematian Bayi (Studi Kasus 38 Kabupaten/Kota di Jawa Timur)," *Forum Statistika dan Komputasi: Indonesia Journal of Statistics*, vol. 17(2), pp. 33-39, 2012.
- [4] N. Delvia, M. and H. Yasin, "*Geographically Weighted Negative Binomial Regression* untuk Menangani Overdispersi pada Jumlah Penduduk Miskin," *Jurnal Gaussian*, vol. 10(4), pp. 532-543, 2021.
- [5] N. M. R. Keswari, I. W. Sumarjaya and N. L. P. Suciptawati, "Perbandingan Regresi Binomial Negatif dan Regresi Generalisasi Poisson dalam Mengatasi Overdispersi (Studi Kasus: Jumlah Tenaga Kerja Usaha Pencetak Genteng di Br. Dukuh, Desa Pejaten)," *E-Jurnal Matematika*, vol. 3(3), pp. 107-115, 2014.
- [6] W. Pratama and S. P. Wulandari, "Pemetaan dan Pemodelan Jumlah Kasus *Tuberculosis* (TBC) di Provinsi Jawa Barat dengan Pendekatan *Geographically Weighted Negative Binomial Regression*," *Jurnal Sains dan Seni ITS*, vol. 4(1), pp. 37-42, 2015.
- [7] I. N. Septiani, "Pemodelan Jumlah Kematian Bayi dengan Pendekatan *Geographically Weighted Negative Poisson Regression* (GWPR) Studi Kasus di Provinsi Kalimantan Barat," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 10(1), pp. 99-108, 2021.
- [8] P. H. Sholikhah and S. , "Model *Geographically Weighted Negative Binomial Regression* pada Kasus Jumlah Rumah Tangga Miskin di Kabupaten Tuban," *Jurnal Sains dan Seni ITS*, vol. 11(2), pp. 235-242, 2022.
- [9] E. Evadiani and P. , "Pemodelan Jumlah Kematian Ibu di Jawa Timur dengan *Geographically Weighted Negative Binomial Regression*," *Jurnal Sains dan Seni POMITS*, vol. 3(2), pp. 182-187, 2014.
- [10] B. W. Y. Priambodo and I. , "Pemetaan Jumlah Property Crime di Provinsi Jawa Timur Menggunakan Metode *Geographically Weighted Negative Binomial Regression* (GWNBR) dan *Geographically Weighted Poisson Regression* (GWPR)," *INFERENSI*, vol. 2(2), pp. 53-62, 2019.
- [11] E. D. Safire and P. , "Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Diabetes Melitus di Jawa Timur Menggunakan GWGPR dan GWNBR," *INFERENSI*, vol. 6(1), pp. 1-14, 2023.
- [12] A. S. Nur Zaina, R. S. Pontoh and B. Tantular, "Pemetaan dan Pemodelan Penyakit TB Paru di Kota Bandung Menggunakan *Geographically Weighted Negative Binomial Regression* (Studi Kasus Dinas Kesehatan Kota Bandung)," *Seminar Nasional Statistika Online (SNSO 2021)*, pp. 67-75, 2021.

EVA SELVIA : Universitas Tanjungpura, Pontianak, evaselviaak@student.untan.ac.id  
NURFITRI IMRO'AH : Universitas Tanjungpura, Pontianak, nurfitriimroah@math.untan.ac.id  
WIRDA ANDANI : Universitas Tanjungpura, Pontianak, wirda.andani@math.untan.ac.id

---