Metode Regresi Poisson Terboboti Geografis pada Pemodelan Data Spasial

Yohana Enggar Setyarini[1], Suyono[2], Widyanti Rahayu[3]
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,Universitas Negeri Jakarta
Jl. Rawamangun Muka Jakarta Timur13320
E-mail: yohanaenggar@gmail.com

ABSTRAK

Data spasial adalah data yang memiliki informasi geografis. Data spasial dapat memiliki pengaruh spasial terhadap variabel terikat dalam bentuk heterogenitas spasial atau dependensi spasial. Oleh karena itu, diperlukan pemodelan spasial yang dapat digunakan untuk menampung pengaruh spasial tersebut. Jika variabel terikat berdistribusi Poisson maka pemodelan spasial yang tepat digunakan adalah model Regresi Poisson Terboboti Geografis (RPTG). Model RPTG merupakan bentuk spasial dari regresi Poisson global. Model RPTG diestimasi menggunakan metode maksimum likelihood dan dilanjutkan dengan metode Newton-Raphson. Model RPTG menghasilkan estimasi parameter yang tidak stasioner atau berbeda-beda untuk setiap wilayah. Model RPTG dalam penelitian ini digunakan untuk memodelkan angka kematian penderita DBD di Jawa Timur tahun 2013. Hasil yang diperoleh menunjukkan bahwa terdapat perbedaan variabel-variabel yang berpengaruh signifikan terhadap variabel terikat antara wilayah yang satu dengan wilayah yang lain. Model RPTG lebih baik dalam memodelkan angka kematian penderita DBD dibanding regresi Poisson global. Model RPTG memiliki nilai AIC=53.205 dan R^2_{dev}=74.16, sedangkan model regresi Poisson global memiliki nilai AIC=59.6301 dan R^2_{dev}=64.83.

Kata kunci : Data Spasial, Pengaruh Spasial, Pemodelan Spasial, Regresi Poisson Global, RPTG.

I. PENDAHULUAN

Perkembangan jumlah penduduk di Jawa Timur cenderung mengalami peningkatan dari tahun ke tahun, seiring dengan bertambahnya jumlah penduduk, maka bertambah juga permasalahan-permasalahan yang terjadi di masyarakat. Permasalahan-permasalahan muncul dari berbagai bidang kehidupan, antara lain: permasalahan di bidang sosial, ekonomi, pendidikan, budaya, kesehatan dan lain sebagainya. Salah satu permasalahan yang terjadi di Jawa Timur dalam bidang kesehatan adalah kasus Demam Berdarah Dengue. Tingginya kasus DBD di Jawa Timur tak jarang menjadi Kejadian Luar Biasa dimana banyak korban meninggal akibat penyakit tersebut. Untuk menekan tingginya kasus kematian penderita DBD, perlu diketahui faktor-faktor apa sajakah yang mempengaruhi angka kematian penderita DBD. Salah satu metode statistika yang dapat digunakan untuk mengetahui apakah terdapat hubungan antara angka kematian akibat DBD dengan faktor-faktor yang diduga memiliki pengaruh terhadap angka kematian penderita DBD adalah analisis regresi. Analisis regresi yang cocok digunakan untuk memodelkan angka kematian penderita DBD adalah regresi Poisson, karena jumlah penderita DBD merupakan data cacah dan merupakan kejadian dengan peluang kejadian yang kecil.

Data kematian penderita DBD diperoleh dari berbagai wilayah Kabupaten atau Kota di Jawa Timur, dengan adanya informasi letak lintang dan bujur dari lokasi data, maka data tersebut disebut data spasial. Data spasial adalah data yang memiliki informasi geografis, misalnya informasi letak bujur dan lintang suatu wilayah. Data spasial memiliki dua elemen penting didalamnya, yaitu informasi geografis (koordinat wilayah), serta atribut (variabel bebas dan variabel terikat) dari wilayah yang diteliti. Data spasial perlu dimodelkan secara tepat, agar tidak menghasilkan kesimpulan yang keliru. Pada pemodelan data spasial perlu diperhatikan adanya pengaruh spasial berupa heterogenitas spasial maupun dependensi spasial yang terdapat pada data spasial yang dapat mempengaruhi variabel terikat.

Pemodelan spasial yang dapat digunakan untuk memodelkan data spasial yang memiliki dependensi spasial maupun heterogenitas spasial adalah model Regresi Poisson Terboboti Geografis (RPTG). Model RPTG merupakan bentuk spasial dari regresi Poisson global. RPTG digunakan pada data dimana variabel terikatnya merupakan data diskrit yang berdistribusi Poisson. Model RPTG menghasilkan estimasi parameter yang tidak stasioner atau berbeda-beda untuk setiap wilayah, sehingga variabel bebas yang signifikan berpengaruh terhadap variabel terikat di suatu wilayah belum tentu signifikan di wilayah yang lain. Model RPTG menampung pengaruh keterkaitan spasial ke dalam bentuk pembobot. Pembobot adalah bentuk kuantifikasi besar pengaruh yang diberikan suatu wilayah terhadap wilayah amatan. Wilayah yang jaraknya lebih dekat dengan titik amatan diasumsikan memiliki pengaruh yang lebih besar dibandingkan dengan wilayah yang jauh dari titik amatan.

Tujuan dari penulisan ini adalah untuk mengkaji model Regresi Poisson Terboboti Geografis, mengetahui variabel-variabel bebas mana saja yang berpengaruh signifikan terhadap variabel terikat pada tiap-tiap wilayah, serta mengetahui keakuratan model Regresi Poisson Terboboti Geografis untuk memodelkan data spasial dibandingkan dengan model regresi Poisson global.

II. LANDASAN TEORI

A. Regresi Linier Berganda

Analisis regresi linier berganda merupakan salah satu metode yang dapat digunakan untuk melihat hubungan antara varibel tak bebas Y dengan lebih dari satu variabel bebas X. Persamaan regresi linier berganda dalam bentuk matriks adalah adalah sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

dimana Y merupakan vektor variabel terikat berukuran $n \times 1$, X merupakan matriks variabel bebas berukuran berukuran $n \times (k+1)$, ε merupakan galat acak yang tidak berkorelasi serta berdistribusi $N(0, \sigma^2)$ berukuran $n \times 1$, dan $\boldsymbol{\beta}$ merupakan parameter yang tidak diketahui nilainya, sehingga perlu dilakukan pendugaan.

B. Distribusi Poisson

Distribusi peluang peubah acak Poisson Y, dinyatakan sebagai berikut:

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

μ menyatakan rata-rata banyaknya sukses yang terjadi per satuan waktu atau daerah, $\mu > 0$ dan $\mathrm{e} = 2.71828\ldots$

Untuk menentukan apakah suatu sampel mengikuti distribusi probabilitas Poisson, dapat digunakan Uji *Kolmogorov-Smirnov*, dengan hipotesis sebagai berikut:

H_0 : sampel berdistribusi Poisson
H_1 : sampel tidak berdistribusi Poisson
dengan statistik uji

$$D_{hitung} = maksimum|F_0(x) - S_N(x)|$$

dimana
$S_N(x)$ = distribusi kumulatif data sampel
$F_0(x)$ = distribusi kumulatif dari distribusi Poisson

Tolak H_0 pada taraf signifikansi α jika $D_{hitung} > D_{(\alpha,N)}$. Nilai $D_{(\alpha,N)}$ diperoleh dari tabel *Kolmogorov-Smirnov*.

C. Regresi Poisson

Regresi Poisson adalah metode statistika yang dapat digunakan untuk menyelidiki hubungan antara variabel terikat yang berupa data diskrit yang berdistribusi Poisson dengan variabel bebas yang bersifat diskrit maupun kontinu. Model regresi Poisson dapat ditulis sebagai berikut:

$$\mu_i = e^{(\boldsymbol{x}_i' \boldsymbol{\beta})}$$

dimana $\boldsymbol{\beta}' = (\beta_0\ \beta_1 \ldots \beta_k)$, dan $\boldsymbol{x}_i' = (1\ x_{i1}\ x_{i2}\ x_{i3} \ldots x_{ik})$

Pendugaan parameter untuk model regresi Poisson digunakan metode maksimum likelihood. Taksiran maksimum likelihood untuk $\boldsymbol{\beta}$ diperoleh dari penyelesaian dari turunan pertama fungsi $l(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i\boldsymbol{x}_i' \ \boldsymbol{\beta} - e^{\boldsymbol{x}_i'\boldsymbol{\beta}} - \ln y_i!)$ terhadap $\boldsymbol{\beta}_j$ dimana $j = 0,1,2, \dots, k$ yang disamakan dengan nol.

D. Pemilihan model terbaik

Pemilihan model terbaik dalam penelitian ini didasarkan pada nilai AIC dan R^2_{dev}

$$AIC = D + 2k$$

dengan $D = 2\sum_{i=1}^{n}(y_i \ln\frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i))$ dan k merupakan jumlah parameter dalam model termasuk konstanta. Nilai koefisien determinasi devians dapat diperoleh dari:

$$R^2_{dev} = 1 - \frac{\sum_{i=1}^{n}(y_i \ln\frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i))}{\sum_{i=1}^{n} \ y_i \ln\frac{y_i}{\bar{y}}}$$

E. Pengaruh Spasial

Data spasial harus dimodelkan secara tepat, jika tidak akan menimbulkan hasil yang keliru, Data spasial dapat memiliki pengaruh spasial terhadap variabel terikat, menurut Anselin(1988) pengaruh spasial terbagi menjadi dua yaitu heterogenitas spasial dan dependensi spasial.

1. Heterogenitas Spasial

Heterogenitas spasial adalah adanya perbedaan karakteristik antara suatu wilayah dengan wilayah lainnya, yang ditandai oleh adanya keunikan dari masing-masing wilayah. Heterogenitas spasial dapat diuji dengan menggunakan statistik uji Breusch-Pagan dengan hipotesis sebagai berikut:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$(tidak terdapat heterogenitas spasial)
$H_1: \sigma_i^2 \neq \sigma_j^2$(terdapat heterogenitas spasial)
Statistik Uji

$$BP = \left(\frac{1}{2}\right) f'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'f$$

Dengan

$$f_i = \left(\frac{y_i - \hat{\mu}_i}{\sigma_i^2} - 1\right)$$

$\hat{\mu}_i = \sigma_i^2 = \exp(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$
Z merupakan matriks berukuran $n \times (k + 1)$ yang berisi variabel bebas yang sudah distandarkan untuk setiap observasi.
Keputusan: Tolak H_0 jika $BP > \chi^2_{\alpha,k}$

2. Dependensi Spasial

Dependensi spasial yaitu keterkaitan antara wilayah yang satu dengan wilayah yang lain, dalam hal ini wilayah yang satu dapat mempengaruhi wilayah yang lain, sehingga perubahan nilai variabel di suatu lokasi dapat mempengaruhi perubahan nilai variabel di lokasi lain yang berdekatan.

Statistik yang dapat digunakan untuk mengetahui apakah terdapat dependensi spasial atau tidak adalah statistik Global Moran's I dengan hipotesis sebagai berikut:

$H_0: I_m = 0$ (tidak terdapat dependensi antar lokasi)
$H_1: I_m = 0$ (terdapat dependensi antar lokasi)
Statistik Uji

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

dengan

$$E(I) = -\frac{1}{n - 1}$$

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

$$Var(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{(n^2-1)S_0^2} - E(I)^2$$

dimana

$\bar{y}$: rata-rata pengamatan di seluruh wilayah

w_{ij} : pembobot spasial yang diperoleh dari *Queen Contiguity*

$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$

$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}\ + w_{ji})^2$

$S_2 = \sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij}\ + \sum_{j=1}^{n} w_{ji}\ )^2$

Pengambilan keputusan H_0 ditolak jika $\left|Z_{hitung}\ \right| > Z_{\frac{\alpha}{2}}$

Jika H_0hal ini menunjukkan adanya keterkaitan antara suatu data yang diperoleh dari suatu lokasi dengan data yang diperoleh di lokasi lain yang berdekatan, yang dipengaruhi oleh faktor letak geografis dimana data tersebut diambil.

III. PEMBAHASAN

A. Regresi Poisson Terboboti Geografis

Regresi Poisson Terboboti Geografis(RPTG) merupakan bentuk spasial dari regresi Poisson global dimana pada RPTG memberi bobot yang berbeda untuk setiap wilayahnya. Perbedaan antara regresi Poisson global dengan RPTG terdapat pada estimasi parameter, model regresi Poisson global memiliki estimasi parameter yang sama untuk setiap lokasi pengamatan sehingga bersifat global, namun model RPTG menghasilkan estimasi parameter yang berbeda untuk setiap wilayahnya, sehingga bersifat lokal. Model RPTG untuk lokasi ke-i dapat ditulis sebagai berikut (Nakaya,2005):

$$\ln \mu_i = \beta_0(u_i, v_i) + \sum_{q=1}^{k} \beta_q(u_i, v_i) x_{iq}$$

Dengan

μ_i = rataan dari variabel terikat y_i pada wilayah ke$-\ i$

x_{iq} = nilai variabel bebas ke-q pada wilayah ke$-\ i$

$(u_i, v_i) =$ titik koordinat wilayah ke – i

$\beta_0(u_i, v_i) =$ konstanta pada wilayah ke – i

$\beta_q(u_i, v_i) =$ parameter regresi dari variabel bebas ke– q pada wilayah ke– i

$i =$ 1,2,…,n

q =1,2,…,n

$k =$ banyaknya variabel bebas

Model RPTG dapat menghasilkan model dengan parameter non-stationer untuk setiap amatan karena RPTG memberikan pembobot pada setiap amatan yang merupakan bentuk kuantifikasi dari pengaruh keterkaitan spasial yang diberikan oleh wilayah lain terhadap suatu lokasi yang diteliti. Diasumsikan wilayah yang dekat dengan titik regresi memiliki pengaruh yang lebih besar dibandingkan dengan wilayah yang jaraknya jauh dari titik regresi. Jarak antara dua wilayah memiliki peranan penting dalam penentuan nilai pembobot. Perhitungan jarak antara dua titik dapat digunakan rumus jarak euclid dengan formula sebagai berikut:

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

dengan (u_i, v_i) dan (u_j, v_j) masing-masing menunjukkan koordinat lintang dan bujur wilayah ke-i dan ke-j. Salah satu cara mendapatkan pembobot adalah dengan menggunakan fungsi kernel.

- Fungsi kernel Gaussian

  Fungsi kernel Gaussian didefinisikan sebagai berikut (Fotheringham, 2002):

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$$

- Fungsi kernel bisquare
  Fungsi kernel bisquare didefinisikan sebagai berikut (Fotheringham, 2002):

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & \text{Jika } d_{ij} < b \\ 0 & \text{jika lainnya} \end{cases}$$

dimana w_{ij} merupakan pembobot wilayah ke-j terhadap wilayah ke-i , d_{ij}merupakan jarak antara titik lokasi yang diteliti (i) dengan titik data yang diperoleh dari lokasi amatan lain (j) dan b merupakan *bandwidth* yang merupakan bilangan nonnegatif yang dianalogikan dengan panjang jari-jari sebuah lingkaran, wilayah-wilayah yang berada di dalam lingkaran, dianggap memiliki pengaruh terhadap titik lokasi yang diteliti. Jika indeks i dan j bernilai sama maka jaraknya nol sehingga nilai pembobotnya akan bernilai satu. Jika indeks i dan j tidak sama maka nilai pembobot w_{ij} akan menurun mengikuti kurva normal sejalan dengan semakin jauhnya jarak i dengan j.
Fungsi pembobot jika dilihat dari *bandwidth* terbagi menjadi dua, yaitu fungsi pembobot dengan *bandwidth* tetap dan fugsi pembobot dengan *bandwidth* adaptive. Pada fungsi kernel tetap sebuah *bandwidth* optimum di tetapkan, kemudian *bandwidth* tersebut diberlakukan untuk setiap titik amatan. Lain halnya fungsi kernel dengan *bandwidth* adaptive, setiap lokasi pengamatan memiliki nilai *bandwidth* yang berbeda satu sama lain, namun memiliki jumlah tetangga terdekat sama.
Berikut adalah contoh fungsi pembobot adaptive bisquare:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b_i}\right)^2\right)^2 & \text{Jika } d_{ij} < b_i \\ 0 & \text{jika lainnya} \end{cases}$$

berikut adalah contoh fungsi pembobot adaptive Gaussian:

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b_i}\right)^2\right)$$

dimana b_i merupakan nilai *bandwidth* yang berbeda-beda untuk setiap wilayah ke-i sesuai dengan banyaknya tetangga terdekat yang telah ditetapkan. Nilai b merupakan jarak titik regresi ke tetangga terdekat pada urutan ke-k, dimana k merupakan jumlah tetangga.

B. Estimasi Parameter model RPTG
Estimasi parameter untuk model RPTG digunakan metode maksimum likelihood. Taksiran maksimum likelihood untuk $\boldsymbol{\beta}(u_i, v_i)$ diperoleh dari penyelesaian dari turunan pertama fungsi

$$l^*\left(\boldsymbol{\beta}(u_{i,}v_i)\right) = \sum_{j=1}^{n}(y_j\boldsymbol{x}_j'\boldsymbol{\beta}\left(u_{i,}v_i\right) - \exp(\boldsymbol{x}_j'\,\boldsymbol{\beta}\left(u_{i,}v_i\right)) - \ln y_j!)w_{ij}\left(u_{i,}v_i\right)$$

terhadap $\boldsymbol{\beta}_q(u_i, v_i)$ dimana $q = 0,1,2, \dots, k$, kemudian disamakan dengan nol. Karena persamaan dari turunan pertama tersebut berbentuk impilsit, maka penyelesaiannya dapat dicari menggunakan pendekatan numerik Newton-Raphson yaitu dengan melakukan iterasi hingga konvergen dengan formula:

$$\widehat{\boldsymbol{\beta}}\left(u_{i,}v_i\right)^{(\boldsymbol{r})} = \widehat{\boldsymbol{\beta}}\left(u_{i,}v_i\right)^{(\boldsymbol{r-1})} + (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{A}^{(\boldsymbol{r-1})}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{A}^{(\boldsymbol{r-1})}\boldsymbol{k}$$

dengan

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & & & & \\ 1 & x_{n1} & x_{12} & \dots & x_{nk} \end{pmatrix}; \quad \mathbf{W}(u_i, v_i) = \begin{pmatrix} w_{i1}(u_i, v_i) & 0 & \cdots & 0 \\ 0 & w_{i2}(u_i, v_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_{in}(u_i, v_i) \end{pmatrix}$$

$$\mathbf{A}(u_i, v_i)^{(r-1)} = \begin{pmatrix} e^{\mathbf{x_1}'\hat{\boldsymbol{\beta}}(u_i,v_i)^{(r-1)}} & 0 & \dots & 0 \\ 0 & e^{\mathbf{x_2}'\hat{\boldsymbol{\beta}}(u_i,v_i)^{(r-1)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \dots \\ 0 & 0 & \dots & e^{\mathbf{x_n}'\hat{\boldsymbol{\beta}}(u_i,v_i)^{(r-1)}} \end{pmatrix}$$

$\boldsymbol{k}$ merupakan vektor berukuran $(n \times 1)$ yang elemen-elemen dalam k dapat ditulis:

$$\mathbf{k_i} = \frac{y_i - \mu_i}{\mu_i}$$

C. Uji Signifikansi Parameter secara Serempak

Pengujian signifikansi parameter secara serempak dilakukan untuk mengetahui apakah terdapat minimal satu variabel bebas yang berpengaruh signifikan terhadap variabel terikat.

Uji hipotesisnya adalah sebagai berikut:

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

$H_1 : \beta_j(u_i, v_i) \neq 0$ dengan $j = 1, 2, \dots, k$ dan $i = 1, 2, 3\dots, n$

dengan statistik uji

$$G = -2(L_1 - L_0)$$

dengan

$L_0 = \sum_{i=1}^{n} (y_i \ln \hat{\mu}_i(u_i, v_i) - \hat{\mu}_i(u_i, v_i) - \ln y_i!)$ dengan $\hat{\mu}_i(u_i, v_i) = \exp(\beta_0(u_i, v_i))$

$L_1 = \sum_{i=1}^{n} (y_i \ln \hat{\mu}_i(u_i, v_i) - \hat{\mu}_i(u_i, v_i) - \ln y_i!)$ dengan $\hat{\mu}_i(u_i, v_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}(u_i, v_i))$

Tolak H_0 jika nilai $G > \chi_{(\alpha,d)}$ dengan d adalah selisih dari parameter efektif yang digunakan.

D. Uji Signifikansi Parameter secara Individu

Pengujian signifikansi parameter secara individu dilakukan untuk mengetahui parameter mana sajakah yang bepengaruh signifikan terhadap variabel terikat.

Uji hipotesisnya adalah sebagai berikut:

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

H_1: $\beta_j(u_i, v_i) \neq 0$ dengan $j = 0,1,2 \dots, k$

Dengan statistik uji

$$t_j(u_i, v_i) = \frac{\beta_j(u_i, v_i)}{SE(\beta_j(u_i, v_i))}$$

dimana

$SE\left(\beta_j(u_i, v_i)\right) = \sqrt{Var(\beta_j(u_i, v_i))}$

Keputusan: Tolak H_0 jika$|t_{hitung}| > t_{v,\frac{\alpha}{2}}$, dimana α adalah tingkat signifikansi dan v adalah derajat bebas yaitu $(n - k - 1)$.

E. Langkah-langkah Penelitian:

1. Menentukan variabel terikat (Y) dan variabel bebas $X_1, X_2, \dots, X_p$, yang akan digunakan dalam penelitian. Kemudian uji sebaran variabel terikat dengan menggunakan kolmogorof smirnof.
2. Melakukan pengujian multikolinieritas antar variabel bebas dengan melihat nilai VIF.
3. Melakukan estimasi parameter model regresi Poisson.
4. Melakukan pengujian parameter regresi Poisson secara serempak menggunakan statistik uji rasio likelihood.
5. Melakukan pengujian parameter regresi Poisson secara parsial dengan menggunakan uji Wald.

6. Melakukan pengujian pengaruh spasial. Pengujian terhadap heterogenitas spasial dapat dilakukan dengan uji Breusch-Pagan, serta uji Global Moran's I untuk menguji dependensi spasial. Jika terdapat pengaruh spasial, maka pemodelan spasial akan dilakukan, yaitu model Regresi Poisson Terboboti Geografis.
7. Berikut adalah tahapan pencarian model spasial Regresi Poisson Terboboti Geografis:
   - Mencari jarak euclid antar wilayah.
   - Mencari model RPTG dengan jumlah tetangga tertentu k, yang mengakibatkan nilai AIC menjadi minimum.
   - Melakukan uji serempak parameter model RPTG
   - Melakukan uji parsial model RPTG
   - Mendapat model terbaik
8. Membandingkan performa antara regresi Poisson dengan regresi Poisson Terboboti Geografis
9. Membuat interpretasi dan kesimpulan berdasarkan model terbaik yang telah diperoleh.

F. Contoh Kasus

Dalam penelitian ini, Model RPTG digunakan untuk memodelkan angka kematian penderita DBD di Jawa Timur tahun 2013. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Kesehatan dan Badan Pusat Statistik. Dalam penelitian ini digunakan data dari 38 Kabupaten/Kota di Jawa Timur. Variabel terikat yang digunakan pada penelitian ini adalah jumlah penderita DBD yang meninggal(Y), variabel bebas yang digunakan dalam penelitian ini yaitu sebanyak 7 variabel. Adapun variabel-variabel bebas tersebut adalah persentase sarana kesehatan baik Puskesmas maupun rumah sakit (X_1), persentase penyuluhan kesehatan yang dilakukan (X_2), persentase masyarakat yang memiliki jaminan kesehatan masyarakat (X_3), ketinggian wilayah dari permukaan laut(X_4), persentase keluarga yang berprilaku hidup bersih dan sehat (X_5), persentase rumah sehat (X_6) dan rata-rata curah hujan per bulan (X_7). Berikut adalah sebaran jumlah kematian penderita DBD di Jawa Timur.

Langkah pertama yang harus dilakukan ialah menguji sebaran dari variabel terikat Y dengan menggunakan uji *Kolmogorov-Smirnov* , hasil yang diperoleh yaitu $0.207 = D_{hitung} < D_{(\alpha,38)} = 0.215$ sehingga dapat disimpulkan variabel terikat Y mengikuti sebaran Poisson.
Selanjutnya, perlu dilakukan uji multikolinieritas dengan melihat nilai VIF dari masing-masing variabel bebas.

Tabel 3.1: Nilai VIF masing-masing variabel bebas

Variabel	X_1	X_2	X_3	X_4	X_5	X_6	X_7
VIF	1.307	1.292	1.637	1.399	1.620	1.353	1.173

Berdasarkan Tabel 3.1 diketahui bahwa semua nilai VIF dari masing-masing variabel bernilai kurang dari 10 artinya tidak terdapat multikolinieritas antar variabel bebas.
Selanjutnya mencari model regresi Poisson global. Model terbaik untuk regresi Poisson global adalah sebagai berikut:

$$\hat{\mu} = e^{0.504X_1 - 0.105X_2 + 0.021X_3 - 0.017X_6}$$

Setiap kenaikan 1% banyaknya penyuluhan kesehatan(X_2) dan variabel lain dijaga tetap(konstan), rata-rata banyaknya kematian penderita DBD akan berkurang sebanyak $e^{\widehat{\beta_2}} = e^{-0.105} = 0.9$ kali. Hal ini juga berlaku untuk variabel lainnya, dimana besarnya penurunan/kenaikan bergantung pada parameter masing-masing variabel. Setelah model regresi Poisson global diperoleh, maka perlu dilakukan pengujian pengaruh spasial.

- **Heterogenitas spasial**

Nilai BP yang diperoleh oleh model regresi Poisson global adalah 25.73 nilai ini lebih besar dari nilai $\chi_{(0.05,4)} = 9.488$, hal ini menunjukkan bahwa terdapat heterogenitas spasial pada model regresi Poisson global.

- **Dependensi spasial**

Uji moran's I dapat digunakan untuk menguji apakah terdapat dependensi spasial atau tidak. Berikut adalah hasil uji morans'I

Tabel 3.2: Hasil uji Moran's I

Variabel	I	Z(I)
Y	-0.07953	-0.399986
X_1	0.11393	1.073774
X_2	0.12657	1.170063
X_3	0.44286	3.579511*
X_6	0.30943	2.563030*

*Signifikan pada $\alpha = 5\%$; $Z_{0.025} = 1.96$

Berdasarkan Tabel 3.2 diketahui bahwa terdapat dependensi spasial pada variabel X_3 dan X_6.

a) Model RPTG

Model regresi Poisson global yang telah diperoleh memiliki pengaruh spasial, maka perlu dilakukan pemodelan spasial yang dapat menampung kedua pengaruh spasial tersebut. Model RPTG akan digunakan untuk mengatasi kedua permasalahan tersebut.

Dalam mekakukan analisis dengan metode RPTG perlu untuk menentukan fungsi pembobot yang akan digunakan, dalam penelitian ini akan digunakan fungsi adaptive bisquare dengan jumlah tetangga 35 karena setelah dilakukan pencarian jumlah tetangga yang menghasilkan AIC paling minimum yaitu dengan 35 tetangga terdekat dengan nilai $AIC = 53.206$.

Langkah berikutnya ialah menghitung jarang euclid untuk masing-masing wilayah, kemudian mencari *bandwidth* untuk masing-masing amatan, selannjutnya hitung nilai pembobot untuk msing-masing wilayah berdasarkan 35 tetangga terdekat.

Estimasi parameter model RPTG menggunakan metode maksimum likelihood yang dilanjutkan dengan metode Newton-Raphson. Sehingga dihasilkan estimasi parameter untuk masing-masing wilayah.

Uji signifikansi parameter model RPTG secara serempak digunakan uji rasio likelihood. Berdasarkan perhitungan diperoleh nilai $101.631 = G > \chi_{(0.05,6)} = 12.592$. Hal ini menunjukkan bahwa minimal terdapat satu variabel bebas yang berpengaruh terhadap model.

Uji signifikansi parameter model RPTG secara individu bertujuan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap Y untuk masing-masing wilayah. Uji signifikansi parameter secara individu dilakukan dengan membagi nilai dugaan dengan simpangan bakunya kemudian dibandingkan dengan nilai t tabel.

Berikut adalah sebaran variabel-variabel bebas yang signifikan berpengaruh terhadap variabel terikat pada masing-masing wilayah di Jawa Timur.

Gambar 3.1 Sebaran variabel-variabel bebas yang sinifikan di tiap wiilayah

Berdasarakan Gambar 3.1 dapat dilihat bahwa variabel-variabel bebas yang signifikan mempengaruhi jumlah kematian penderita DBD di Jawa Timur berbeda-beda antara wilayah yang satu dengan wilayah lainnya. Variabel-variabel bebas yang signifikan terhadap Y tersebut cenderung mengelompok dimana wilayah yang berdekatan memiliki variabel signifikan yang sama. Hal ini diduga karena adanya kesamaan karakteristik antar wilayah yang berdekatan, baik itu keadaan geografis, sosial, kebudayaan, maupun gaya hidup. Faktor yang signifikan di Kabupaten Jember adalah $X_1, X_{2,}, X_6$maka diperoleh model RPTG untuk wilayah ini adalah $\hat{\mu} = e^{0.5325X_1 - 0.1264X_2 - 0.0210X_6}$. Sedangkan model untuk Kabupaten sampang adalah $\hat{\mu} = e^{0.4514X_1 - 0.0147X_6}$. Model yang diperoleh oleh masing-masing daerah berbeda-beda. Terdapat lima Kabupaten/Kota di Jawa Timur dengan variabel-variabel bebas yang signifikan mempengaruhi angka kematian DBD meliputi persentase sarana kesehatan(X_1), persentase penyuluhan kesehatan yangdilakukan(X_2) dan persentase rumah sehat(X_6). Terdapat lima Kabupaten/Kota di Jawa Timur dengan variabel-variabel bebas yang signifikan mempengaruhi angka kematian penderita DBD meliputi persentase sarana kesehatan(X_1), dan persentase rumah sehat(X_6). Serta terdapat 28 Kabupaten/Kota di Jawa Timur, dengan variabel-variabel yang signifikan mempengaruhi angka kematian penderita DBD meliputi persentase sarana kesehatan(X_1), persentase penyuluhan kesehatan yang dilakukan(X_2) , persentase masyarakat yang memiliki jaminan kesehatan masyarakat(X_3) dan persentase rumah sehat(X_6).

b) Perbandingan model regresi Poisson global dengan model RPTG

Perbandingan model regresi Poisson global dengan model RPTG dilakukan untuk melihat keakuratan model lokal RPTG jika dibandingkan dengan model regresi Poisson global. Perbandingan dilakukan dengan membandingkan nilai AIC dan dari kedua metode.

Tabel 3.3 : Nilai Devians, AIC dan R^2_{dev}regresi Poisson global dan RPTG

Model	Devians	AIC	R^2_{dev}
Regresi Poisson global	49.630	59.6301	64.83%
Regresi Poisson Terboboti Geografis	36.462	53.205	74.16%

Berdasarkan Tabel 3.3 diketahui bahwa model RPTG memiliki nilai AIC yang lebih kecil dari regresi Poisson global, serta nilai R^2_{dev} yang lebih besar dari regresi Poisson global, Hal ini menunjukkan bahwa model RPTG lebih sesuai digunakan untuk memodelkan kasus kematian penderita DBD di Jawa Timur dibandingkan dengan model regresi Poisson global. Model RPTG dapat memodelkan data spasial dengan baik karena model RPTG memperhitungkan pengaruh spasial, baik heterogenitas spasial, maupun dependensi spasial.

IV. PENUTUP

A. Kesimpulan

1. Model RPTG untuk lokasi ke-i dapat ditulis sebagai berikut (Nakaya,2005):

$$\ln \mu_i = \beta_0(u_i, v_i) + \sum_{q=1}^{k} \beta_q(u_i, v_i) x_{iq}$$

Dengan
μ_i = rataan dari variabel terikat y_i pada wilayah ke$-\ i$,
x_{iq} = nilai variabel bebas ke-q pada wilayah ke$-\ i$
(u_i, v_i) = titik koordinat wilayah ke – i
$\beta_0(u_i, v_i)$ = konstanta pada wilayah ke – i
$\beta_q(u_i, v_i)$ = parameter regresi dari variabel bebas ke– q pada wilayah ke– i
i = 1,2,…,n
q =1,2,…,n
k = banyaknya variabel bebas

2. Terdapat perbedaan antara variabel bebas yang berpengaruh signifikan terhadap variabel terikat antara wilayah yang satu dengan wilayah yang lain. Berdasarkan hasil yang diperoleh dari pemodelan angka kematian penderita DBD di Jawa Timur dengan RPTG, diperoleh lima wilayah dimana variabel yang berpengaruh signifikan terhadap variabel terikat berupa $X_1, X_{2,} X_6$, terdapat lima wilayah dengan faktor yang signifikan meliputi X_1, X_6, serta terdapat 28 wilayah dengan faktor yang signifikan meliputi $X_1, X_{2,} X_3, X_6$.

3. Model RPTG memiliki nilai AIC yang lebih kecil dari regresi Poisson global, serta nilai R^2_{dev} yang lebih besar dari regresi Poisson global, hal ini menunjukkan bahwa model RPTG lebih sesuai digunakan untuk memodelkan kasus kematian penderita DBD di Jawa Timur dibandingkan

dengan model regresi Poisson global. Nilai AIC model regresi Poisson global adalah 59.6301 serta $R_{dev}^2 = 64.83\%$. Sementara Nilai AIC model regresi Poisson global adalah 53.205 serta $R_{dev}^2 = 74.16\%$.

B. Saran

Penelitian ini dapat dilanjutkan dengan metode Regresi Poisson Terboboti Geografis semiparametrik. Selain itu jika regresi Poisson mengalami overdispersi, metode Regresi Negatif Binomial Terboboti Geografis dapat digunakan.

Daftar Pustaka

Agresti, Alan. 2002. *Categorical Data Analysis*. Edisi Kedua. New York: John Wiley & Sons, Inc.

Astuti, A.D. 2014. Partial Least Square (PLS) dan Principal Component Regression (PCR) untuk Regresi Linier dengan Multikolinieritas pada Kasus Indeks Pembangunan Manusia di Kabupaten Gunung Kidul. *Skripsi*, Universitas Negeri Yogyakarta.

Amelia, Rahmi dan Purhadi. 2012. "Pemodelan Jumlah Balita Gizi Buruk di Jawa Timur dengan Geographically Weighted Poisson Regression". *Jurnal Sains dan Seni ITS, Vol.1, No.1*.

Anselin, Luc. 1988. *Spatial Ekonometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.

Aulele, Salmon N. 2011. "Model Geographically Weighted Poisson Regression dengan Pembobot Fungsi Kernel Gauss". *Jurnal Barekeng, Vol.5 No.2 Hal.25-30*.

Badan Pusat Statistik. 2013. *Jawa Timur Dalam Angka 2013*. Surabaya: BPS Provinsi Jawa Timur.

BPS. 2013. *Pengembangan Model Sosial Ekonomi: Penggunaan Model GWR untuk Analysis Data Spasial dan Ekonomi*. Jakarta: Badan Pusat Statistik.

Cameron, A.C. dan Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge

Dinkes. 2014. *Profil Kesehatan Jawa Timur Tahun 2013*. Dinas Kesehatan Provinsi Jawa Timur.

Dobson, A.J. 2002. *An Introduction to Generalized Linier Models: Chapmann & Hall*. USA.

Draper, N.R. dan Harry Smith. 1998. *Applied Regression Analysis, Three Edition*. New York: John

Fischer, Manfred M dan Jinfeng Wang. 2011. *Spatial Data Analysis: Models, Methods, and techniques*. New York: Springer.

Fotheringham, A., Chris Brunsdon, dan Martin Charlton. 2002. *Geographically Weighted Regression The Analysis Of Spatially Varying Relationships*. UK: John Wiley \& Sons, LTD.

Mariane, T.D. 2011. Pemodelan Kasus Demam Berdarah Dengue(DBD) di Jawa Timur dengan model Poisson dan Binomial Negatif. *Tesis*, Institut Pertanian Bogor.

Nakaya, T., \textit{et al}. 2005. "Geographically Weighted Poisson Regression for Disease Association Mapping". *Statistic in Medicine Journal, Volume 24 Issue 17 pages 2695-2717*.

Sembiring, R.K. 1995. Analisis Regresi. Bandung: Penerbit ITB

Sirait, Timbang, "Overdispersi Karena Kesalahan Spesifikasi Model dan Cara Mengatasinya", *Prosiding Seminar Nasional Sains dan Pendidikan Sains IX, Fakultas Sains dan Matematika, UKSW, Salatiga, 21 Juni 2014, Vol 5, No.1, ISSN: 2087-0922.*

Walpole, Ronald E dan Raymond H. Myers. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan edisi ke-4*. Bandung: ITB.