

Comparing of Car-Bym, Generalized Poisson, and Negative Binomial on Tuberculosis Data in Banyumas Districts*

Pembandingan Model Car-Bym, Generalized Poisson, dan Binomial Negatif pada Data Tuberkolosis di Kabupaten Banyumas

Jajang¹, Budi Pratikno², and Mashuri³

^{1,2,3}Universitas Jenderal Soedirman, Indonesia
‡corresponding author: rzjajang@yahoo.com

Copyright © 2021 Jajang, Budi Pratikno, and Mashuri. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In 2019 the number of people with TB (Tuberculosis) in Banyumas, Central Java, is high (1,910 people have been detected with TB). The number of people infected Tuberculosis (TB) in Banyumas is the count data and it is also the area data. In modeling, the parameter estimation and characteristic of the data need to be considered. Here, we studied comparing Generalized Poisson (GP), negative binomial (NB), and Poisson and CAR.BYM model for TB cases in Banyumas. Here, we use two methods for parameter estimation, maximum likelihood estimation (MLE) and Bayes. The MLE is used for GP and NB models, whereas Bayes is used for Poisson and CAR-BYM. The results showed that Poisson model detected overdispersion where deviance value is 67.38 for 22 degrees of freedom. Therefore, ratio of deviance to degrees of freedom is 3.06 (>1). This indicates that there was overdispersion. The following GP, NB, Poisson-Bayes and CAR-BYM are used to modeling TB data in Banyumas and we compare their RMSE. With refer to RMES criteria, we found that CAR-BYM is the best model for modeling TB in Banyumas because its RMSE is smallest.

Keywords: CAR-BYM, generalized poisson, negative binomial, TB.

* Received: Des 2020; Reviewed: Mar 2021; Published: Mar 2021

1. Pendahuluan

Kasus Tuberculosis (TB) merupakan masalah diberbagai wilayah. Di Kabupaten Banyumas Jawa Tengah pada tahun 2019 kasus TB masih tergolong tinggi yaitu sekitar banyak 1.910 kasus. Penyakit TB disebabkan oleh kuman/bakteri *Mycobacterium tuberculosis*. Kuman ini menyebar melalui udara ketika penderita batuk, bicara, tertawa dan atau bersin. Persebaran secara global sebuah penyakit infeksi yang tidak terpola dapat mengancam kesejahteraan manusia, keluarga dan bahkan kesehatan margastwa lainnya. Pengenalan pola persebaran penyakit menjadi penting untuk manajemen, pengendalian penyakit, dan pengambilan kebijakan yang diperlukan (Iddrisu & Amoako, 2016).

Jumlah kasus yang kena penyakit termasuk kedalam tipe data cacah (*count data*) yang diasumsikan mengikuti distribusi Poisson. Model regresi yang relevan adalah model Poisson yang menggunakan *link function log*. Pada model regresi Poisson mengasumsikan bahwa mean dan variansinya sama atau disebut sebagai asumsi *equidispersion*. Penyimpangan dari *equidispersion* baik itu *overdispersion* atau *underdispersion* dapat disebabkan oleh berbagai hal, salah satu penyebab *overdispersi* adalah adanya kluster-kluster (McCullagh & Nelder, 1989). Kondisi *overdispersion* tidak jarang muncul pada data lapangan dan oleh karena itu penggunaan model regresi Poisson tidak relevan lagi. Beberapa model alternatif untuk pemecahan terhadap permasalahan ini antara lain adalah model *generalized* Poisson (GP) dan binomial negatif (*negative binomial*, NB), dan Model Poisson Log Normal. Model Poisson-Log Normal diturunkan dari campuran distribusi Poisson dengan log normal. Model ini mengasumsikan bahwa parameter pada distribusi Poisson mempunyai tingkat heteroskedastisitas yang signifikan. Dalam kasus terjadinya *overdispersi*, kondisi heteroskedastisitas merupakan hal yang mungkin. Oleh karenanya maka perlu model yang dapat mengakomodasi adanya kasus ini. Banyak model yang dapat digunakan untuk mengatasi terjadinya masalah heteroskedastisitas, satu diantaranya adalah model *Conditional Autoregressive* (CAR).

Model CAR sering digunakan untuk menggambarkan variasi spasial kuantitas yang menarik dalam bentuk ringkasan atau agregat atas sub wilayah atau area. Model CAR telah banyak digunakan dalam berbagai bidang seperti bidang, seperti demografi, ekonomi, epidemiologi dan geografi (De Oliveira, 2012). Model CAR juga sering digunakan untuk menganalisa data spasial yang *irregular* (Kyung & Ghosh, 2014). Salah satu tipe model CAR dalam memodelkan data cacah adalah model log normal Poisson (Besag & Kooperberg, 1995). Dalam model CAR Besag-York -Mollie (CAR-BYM) mencakup komponen *Intrinsic* CAR untuk pemulusan spasial dan komponen pengaruh acak untuk efek heterogenitas non spasial. Model CAR-BYM mempunyai kelebihan karena dapat menangkap efek spasial dan non spasial. Namun demikian tentu masalah estimasi parameter menjadi hal yang perlu diperhatikan dalam model ini. Efek distribusi *conditional* akan berdampak pada perlunya pemilihan metode estimasi yang sesuai. Beberapa metode estimasi Bayes dan model CAR-BYM telah dikaji diantaranya adalah oleh (Aswi et al., 2020; Obaromi, 2019).

Metode estimasi bayes eksak untuk mengestimasi parameter melibatkan distribusi prior pada parameter-parameter model. Dari distribusi prior ini selanjutnya dicari distribusi posterior menggunakan konsep peluang bersyarat. Untuk menentukan distribusi posterior dari parameteranya tentu akan menjadi kesulitan tersendiri ketika

melibatkan banyak parameter. Oleh karena itu metode bayes melalui *Markov Chain Monte Carlo* (MCMC) merupakan pilihan yang realistis. Metode estimasi bayes dengan (*Bayesian estimation*) MCMC merupakan metode yang handal dalam menangani masalah model CAR. Metode MCMC telah banyak dikaji oleh para peneliti di bidang medis, antara lain (Han & Lee, 2013; Hendricks & Neumann, 2020; MacNab et al., 2004; Srinivasan & Venkatesan, 2014; Stern & Cressie, 2000). Berkaitan dengan beberapa hal sebagaimana telah dipaparkan di atas maka kajian terkait dengan masalah model CAR seperti manfaat yang sangat luas dalam berbagai bidang, dan metode estimasi parameter, maka perlu kajian dan uji performa dari model. Untuk itu penelitian ini bertujuan untuk membandingkan beberapa model data cacah yang mengalami overdispersi dengan menggunakan model *generalized Poisson* (GP), *negative binomial* (NB), dan CAR-BYM.

2. Metodologi

2.1 Bahan dan Data

Data yang digunakan adalah data sekunder diperoleh dari Dinas Kesehatan (Dinkes) Kabupaten Banyumas dan data Badan Pusat Statistik (BPS) Kabupaten Banyumas. Data yang digunakan dalam penelitian adalah data TB di Kabupaten Banyumas tahun 2019. Variabel-variabel penelitian yang digunakan adalah banyaknya orang yang kena TB. Sedangkan beberapa variabel yang dianggap mempengaruhi jumlah TB adalah banyaknya rumah sehat, banyaknya yang kena HIV, banyaknya yang kena diabetes, total jumlah penduduk dan persentase laki-laki.

2.2 Metode Penelitian

Sebelum ke Langkah-langkah teknis, awal metode penelitian yang digunakan adalah dengan mengulas data area dan matriks pembobot spasial. Bahasan kedua hal ini untuk menghubungkan antara data kasus TB dengan model yang akan digunakan. Berikutnya adalah masuk ke langkah-langkah teknis. Tahapan penelitian yang dilakukan terdiri dari lima. Pertama membentuk fungsi likelihood dari fungsi massa peluang Poisson, NB dan GP. Kedua, menurunkan fungsi likelihood untuk mendapatkan estimator. Ketiga, memodelkan data TB menggunakan *software* R 4.0.0. Keempat mengevaluasi model dengan cara membandingkan tingkat akurasi menggunakan kriteria RMSE dan plot grafik.

2.2.1 Data area dan matriks pembobot spasial

Tipe data area mempunyai karakteristik dimana D yang bersifat tetap dipartisi ke dalam jumlah berhingga unit area yang saling terpisah (Banerjee et al., 2014; Cressie, 1993; Gaetan & Guyon, 2010). Pemodelan pada data spasial melibatkan matriks, yang disebut matriks pembobot spasial. Matriks pembobot spasial ($W=\{w_{ij}\}, i,j=1,2,\dots,n$) adalah *sebuah* matriks tak negatif berukuran $n \times n$. Matriks pembobot spasial yang sering digunakan bernilai 0 dan 1 yang menggambarkan keeratan antar wilayah. Misalkan $W=\{w_{ij}\}, i,j=1,2,\dots,n$, w_{ij} adalah elemen i dan j , dimana $w_{ij}=1$ jika antar spasial saling berdekatan atau mempunyai batas bersama, dan selainnya bernilai nol, dan untuk $w_{ii}=0$.

2.2.2 Model Poisson, GP dan NB

Model Poisson adalah model yang digunakan untuk data respon cacah yang berdistribusi Poisson. Misalkan $Y_i \sim Poi(\mu_i)$, maka fungsi massa peluang (pmf) dari Y_i adalah sebagai berikut

$$p(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0,1,2,\dots, E(Y_i) = \mu_i, Var(Y_i) = \mu_i, \quad (1)$$

Distribution Poisson termasuk keluarga eksponensial dan dapat dinyatakan sebagai

$$p(y_i) = \frac{1}{y_i!} e^{[y_i \ln(\mu_i) - \mu_i]}, \text{ sehingga link function model adalah } \log, \eta_i = \ln(\mu_i) \Leftrightarrow \mu_i = e^{\eta_i}.$$

Jika terdapat $k (X_j, j=1,2,\dots,k)$ variabel prediktor, maka fungsi $g, g(\mu_i) = \ln(\mu_i) = X_i' \beta$ untuk estimasi Parameter Model Poisson dengan MLE, tahap awal yang dilakukan adalah membentuk fungsi *likelihood*. Dalam model Poisson untuk memaksimumkan fungsi *likelihood* berdasarkan parameter-parameternya raltif sulit. Oleh karenanya fungsi *likelihood* ditransformasi ke *log likelihood*. Transformasi logaritma ini tidak akan mengubah kerelevanan memaksimumkan antara fungsi *likelihood* dan *log likelihood* karena fungsi logaritma bersifat monoton naik. Fungsi *likelihood* merupakan fungsi

Bersama atau fungsi gabungan dari fungsi massa peluang $p(y_i) = \frac{1}{y_i!} e^{[y_i \ln(\mu_i) - \mu_i]}$ dengan

$\eta_i = \ln(\mu_i) \Leftrightarrow \mu_i = e^{\eta_i}$. Dari sini dapat dibentuk *log likelihood*nya. Untuk memaksimumkan fungsi *loglikelihood*, dapat dilakukan dengan mencari titik stasioner, yaitu nilai nilai-nilai parameter sedemikian sehingga nilai fungsi *likelihood* maksimum. Turunan-turunan fungsinya adalah

$$\frac{\partial \ell(\beta; y)}{\partial \beta_0} = \sum_{i=1}^n [y_i - \exp(X_i' \beta)], \text{ dan } \frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum_{i=1}^n [y_i - \exp(X_i' \beta)] x_{ij}; j = 1, 2, \dots, k. \quad (2)$$

Model regresi NB dapat digunakan untuk memodelkan data Poisson yang mengalami *overdispersion* ataupun *underdispersion*. Fungsi massa peluang binomial negative adalah

$$f(y, \mu, m) = \exp \left[y \ln \left(\frac{m\mu}{1+m\mu} \right) + \frac{1}{m} \ln \left(\frac{1}{1+m\mu} \right) + \ln \frac{\Gamma(y + \frac{1}{m})}{y! \Gamma(\frac{1}{m})} \right]. \quad (3)$$

Selanjutnya, dengan membuat fungsi *likelihood* dan *log likelihood*. Fungsi *loglikelihood* distribusi binomial negative adalah

$$\ln L(\beta, m) = \ln \prod_{i=1}^n \exp \left[y_i \ln \left(\frac{m\mu_i}{1+m\mu_i} \right) + \frac{1}{m} \ln \left(\frac{1}{1+m\mu_i} \right) + \ln \frac{\Gamma(y_i + \frac{1}{m})}{y_i! \Gamma(\frac{1}{m})} \right] \quad (4)$$

maka akan diperoleh turunan pertama fungsi *log likelihood* model binomial negatif terhadap Turunan-turunan tersebut adalah

$$\frac{\partial \ell(\beta, m)}{\partial \beta_0} = \sum_{i=1}^n \frac{y_i - \exp(x_i' \beta)}{1 + m \cdot \exp(x_i' \beta)}; \text{ dan } \frac{\partial \ell(\beta, m)}{\partial \beta_j} = \sum_{i=1}^n \frac{x_{ij} (y_i - \exp(x_i' \beta))}{1 + m \cdot \exp(x_i' \beta)}; j = 1, 2, \dots, k. \quad (5)$$

Selanjutnya turunan pertama terhadap parameter disperse, m :

$$\frac{\partial \ell(\beta, m)}{\partial m} = \sum_{i=1}^n \left[\frac{1}{m^2} \left(\ln(1 + m \cdot \exp(x'_i \beta)) + \frac{m(y_i - \exp(x'_i \beta))}{1 + m \cdot \exp(x'_i \beta)} \right) + \psi \left(y_i + \frac{1}{m} \right) - \psi \left(\frac{1}{m} \right) \right]. \quad (6)$$

Disamping model binomial negative, model lain yang dapat digunakan untuk mengatasi masalah penyimpangan terhadap equidispersi adalah model Regresi *Generalized Poisson* (GP). Model GP dapat digunakan baik dalam keadaan *underdispersion*, *equidispersion*, ataupun *overdispersion*. Fungsi massa peluang GP adalah

$$f(y, \mu, m) = \left(\frac{\mu_i}{1 + m\mu_i} \right)^{y_i} \frac{(1 + m y_i)^{y_i - 1}}{y_i!} \cdot \exp \left[\frac{-\mu_i (1 + m y_i)}{1 + m\mu_i} \right] \quad (7)$$

Sama halnya dengan pada distribusi binomial negative, untuk mengestimasi parameter model pada model GP adalah membentuk likelihood dan loglikelihood. fungsi *log likelihood* model GP.

$$\ell(\beta, m) = \sum_{i=1}^n y_i \ln \left(\frac{\mu_i}{1 + m\mu_i} \right) + (y_i - 1) \ln(1 + m y_i) - \left[\frac{-\mu_i (1 + m y_i)}{1 + m\mu_i} \right] - \ln \Gamma(y_i + 1); \quad (8)$$

Turunan pertama terhadap β_0 dan β_j :

$$\frac{\partial \ell(\beta, m)}{\partial \beta_0} = \sum_{i=1}^n \frac{(y_i - \exp(x'_i \beta))}{(1 + m \cdot \exp(x'_i \beta))^2}, \text{ dan } \frac{\partial \ell(\beta, m)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \exp(x'_i \beta)) x_{ji}}{(1 + m \cdot \exp(x'_i \beta))^2}; \quad j = 1, 2, \dots, k. \quad (9)$$

Turunan pertama terhadap m :

$$\frac{\partial \ell(\beta, m)}{\partial m} = \sum_{i=1}^n \left\{ -\frac{y_i \cdot \exp(x'_i \beta)}{1 + m \cdot \exp(x'_i \beta)} + \frac{y_i (y_i - 1)}{1 + m y_i} - \frac{\exp(x'_i \beta) \cdot y_i - \exp(x'_i \beta)}{(1 + m \cdot \exp(x'_i \beta))^2} \right\}. \quad (10)$$

Estimasi parameter untuk regresi Poisson, binomial negatif, dan *generalized Poisson* menggunakan iterasi Newton-Raphson. Estimasi parameter dengan iterasi Newton-Raphson menggunakan nilai turunan pertama dan turunan kedua dari fungsi *log-likelihood* sebagai dasar estimasi parameter, yaitu:

$$\beta_{j+1} = \beta_j - \frac{\partial \ell(\beta_j)}{\partial^2 \ell(\beta_j)}; \quad j = 0, 1, 2, \dots, k. \quad (11)$$

dengan β_{j+1} dan β_j adalah parameter untuk $\beta = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$ pada iterasi ke $j+1$ dan j untuk iterasi Newton-Raphson.

2.2.3 Model Poisson Spasial Estimasi Bayesian

Metode Estimasi Bayesian ditentukan oleh distribusi posterior. Misalkan y_i , untuk $i=1, 2, \dots, n$ adalah sampel acak yang dari probability density function $P(y|\theta)$, dengan θ parameter $\theta = (\theta_1, \dots, \theta_p)$ maka fungsi densitas bersama adalah $P(y|\theta) = \prod_{i=1}^n P(y_i|\theta)$. Ambil untuk satu parameter θ , dengan distribusi priornya $p(\theta)$, maka distribusi posteriornya

$$P(\theta|y) = \frac{P(y,\theta)}{p(y)} = \frac{P(y|\theta)P(\theta)}{p(y)} \tag{12}$$

Estimator dengan metode Bayes adalah $\hat{\theta} = E(\theta|y)$ (Ibrahim et al., 2001).

2.2.4 Model Conditional Autoregressive (CAR)

Model CAR didasari oleh model Poisson lognormal. Model Poisson-Lognormal diturunkan dari campuran distribusi Poisson dengan lognormal. Model ini mengasumsikan bahwa parameter Poisson mempunyai tingkat heteroskedastisitas signifikan, sehingga dimodelkan kembali. Misalkan Y_i variable acak yang mengikuti distribusi Poisson, $Y_i \sim POI(\mu_i)$, $\mu_i = E_i \theta_i$ dan $\theta_i = \exp(\eta_i)$, $\eta_i = X' \beta + v_i$, v_i mempunyai struktur CAR. Model log normal didefinisikan sebagai

$$\log(\mu_i) = \log(E_i) + x_i' \beta + v_i \tag{13}$$

dengan μ_i adalah mean Y_i , $\log(E_i)$ adalah *offset*, x_i' adalah vektor variabel prediktor, β vektor parameter, v_i error yang mengikuti CAR. Fungsi likelihood dari θ_i adalah $l(\beta, v) = \prod_{i=1}^n \frac{e^{-E_i \theta_i (E_i \theta_i)^{y_i}}}{y_i !} = P(y, E, \theta | \beta, v)$. Ambil distribusi prior untuk β dan v ,

$vp(\beta) = \left(\frac{1}{2\pi}\right)^{\frac{P}{2}} \left(\frac{1}{\tau_\beta}\right)^P \exp\left(-\frac{1}{2} \sum_{h=0}^P \frac{\beta_h^2}{\tau_\beta^2}\right)$ dan prior untuk CAR efek acak adalah $[v_i | v_{i \neq j}, \tau_v^2] \sim CAR(0, \tau_v^2) \sim N\left(\sum_{i \neq j}^n \frac{w_{ij} \cdot v_j}{w_{ij}}, \frac{\tau_v^2}{w_{ij}}\right)$. Dengan menggunakan (12) diperoleh

$$p(\beta, v, \tau_\beta^2, \tau_v^2 | y, E, \theta) = \prod_{i=1}^n \frac{e^{-E_i \theta_i (E_i \theta_i)^{y_i}}}{y_i !} x \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{P}{2}} \left(\frac{1}{\tau_\beta}\right)^P \exp\left(-\frac{1}{2} \sum_{h=0}^P \frac{\beta_h^2}{\tau_\beta^2}\right) x \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=0}^n \left(\frac{v_i - \sum_{i \neq j}^n \frac{w_{ij} \cdot v_j}{w_{ij}}}{\tau_\beta}\right)^2\right)$$

2.2.5 Model Conditional Autoregressive –BYM (CAR-BYM)

Model Conditional Autoregressive –BYM (CAR-BYM) mencakup komponen ICAR untuk penghalusan spasial dan komponen efek acak biasa untuk heterogenitas non-spasial. Model BYM dispesifikasi sebagai berikut:

$$\eta_i = \mu + x'_i \beta + \phi_i + \theta_i, \quad i = 1, 2, \dots, n, \tag{14}$$

dengan

= vector koefisien regresi bersifat tetap

ϕ_i = komponen ICAR

μ = rata-rata risk level

θ_i = efek acak komponen non-spatial heterogenitas

Model BYM (14) mengakomodasi komponen spasial (ϕ) dan komponen non spasial (θ) untuk efek pengaruh overdispersi yang tidak bisa dipecahkan dengan regresi

Poisson. Estimasi parameter dengan metode Bayes merupakan pilihan dalam hal ini karena selain relatif lebih longgar dalam asumsi, metode ini bisa digunakan untuk masalah distribusi bersyarat seperti yang ada dalam model CAR.

2.2.6 Markov Chain Monte Carlo

Metode Markov Chain Monte Carlo (MCMC) menggunakan penarikan sampel Gibbs (*Gibb sampler*) untuk mengestimasi parameter. Metode Gibb sampler adalah: Untuk $t=1,2,\dots,T$

- Tahap 1. Ambil θ_1^t dari $p(\theta_1 | \theta_2^{t-1}, \theta_3^{t-1}, \dots, \theta_k^{t-1}, y)$
- ...
- Tahap k. Ambil θ_k^t dari $p(\theta_k | \theta_1^{t-1}, \theta_2^{t-1}, \dots, \theta_{k-1}^{t-1}, y)$

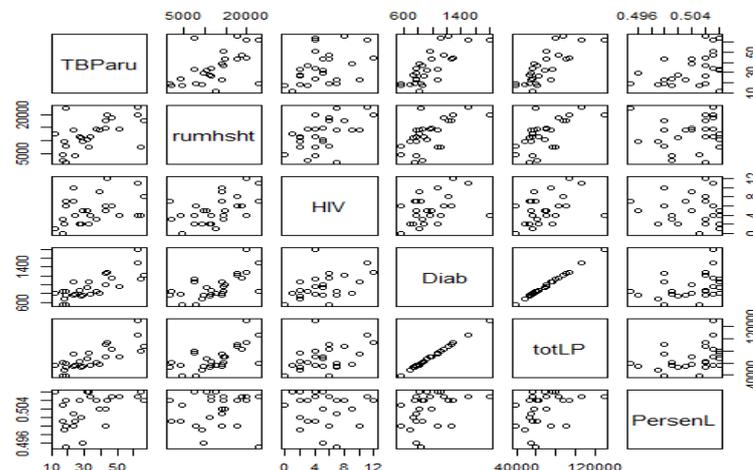
Misalkan kondisi stabil ketika terjadi setelah iterasi t_0 , maka estimasi parameter θ_i

$$\hat{\theta}_i = \hat{E}(\theta_i | y) = \frac{1}{T - t_0} \sum_{t=t_0+1}^T \theta_i^t$$

Karena solusi eksak Bayes relatif lebih sulit karena pencarian integral dari fungsi distribusi posteriornya tidak sederhana, maka untuk selanjutnya metode estimasi parameter Bayes menggunakan Markov Chain Monte Carlo (MCMC).

3. Hasil dan Pembahasan

Tahapan eksplorasi merupakan tahap awal yang perlu dilakukan sebelum melakukan analisis lebih lanjut. Pada tahap ini akan dilakukan dengan *scatterplot* antara variabel respon dengan variabel prediktor. Hasil *scatterplot* antara variabel respon (jumlah TB) dan beberapa variabel prediktornya daisajikan pada Gambar 1.

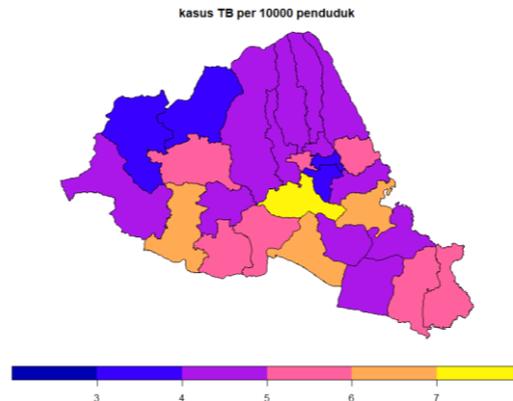


Gambar 1: *Scatterplot* antar variabel penelitian

Berdasarkan *scatterplot*, variabel prediktor jumlah orang diabetes berkorelasi dengan total penduduk, sehingga diambil salah satu variabel, jumlah diabetes. Sementara setiap variabel prediktor mempunyai pola hubungan dengan variabel respon (Gambar 1).

3.1 Estimasi proporsi Bayes Empiris

Analisis deskripsi pemetaan proporsi dalam hal ini adalah dengan menggunakan metode *Empirical Bayes* dengan bantuan *software R software 4.0.0*. Jumlah populasi yang terkadang cukup jauh membutuhkan metode metode Bayes empiris, menggunakan Poisson.

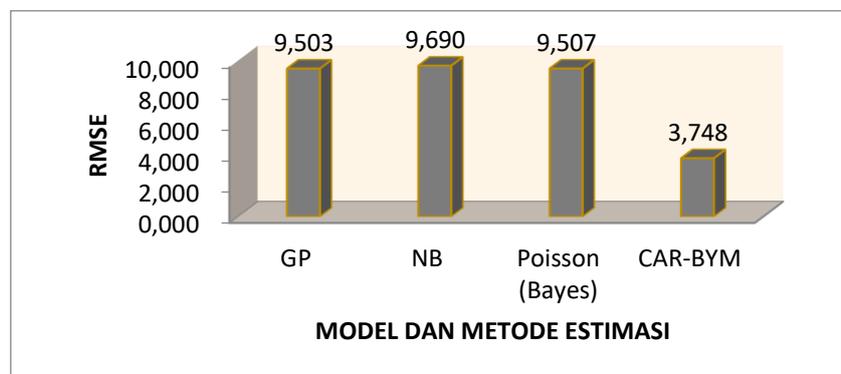


Gambar 2: Eksplorasi *Empirical Bayes estimation* di Kabupaten Banyumas

Jumlah kasus per seribu pendudukan terlihat bahwa dominasi berada di angka 4-5 kasus per 10000 penduduk. Sedangkan jumlah kasus yang berada di atas 7 kasus per 10000 orang penduduk (Kecamatan Patikraja) hanya ada satu kecamatan.

3.2 Implementasi Model

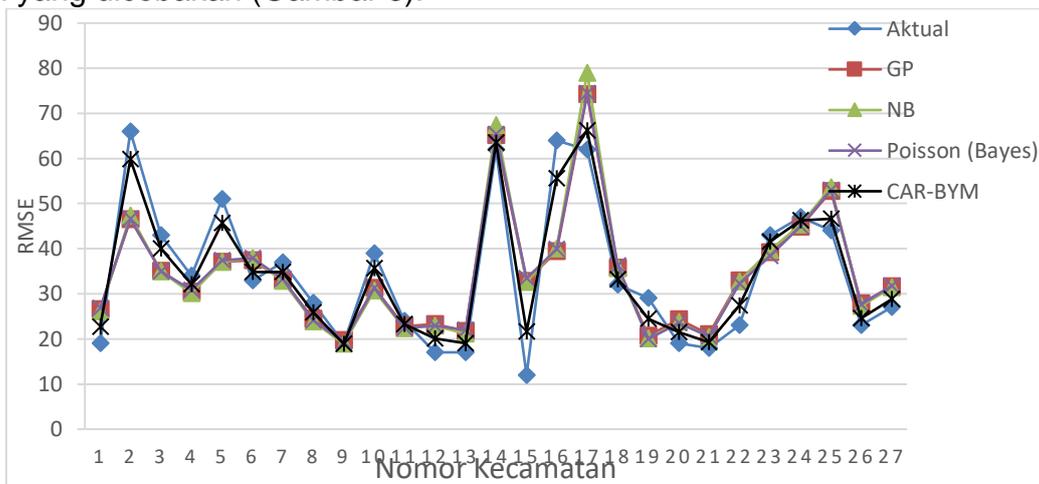
Tahapan awal adalah pengujian overdispersi model Poisson, dari sini diperoleh diperoleh *deviance* 67.37833 dan *df* 22, sehingga $dev/df = 3.062652 (>1)$ terjadi overdispersi. Selanjutnya digunakan model GP dn NB untuk menyelidiki dampak overdispersi terhadap hasil akurasi model. Selain itu dibandingkan pula dengan model Poisson metode Bayes dan model CAR-BYM. Hasil pembandingannya RMSE dari keempat model dapat dilihat pada Gambar 3. Gambar 3 merupakan diagram batang yang merepresentasikan hubungan antara RMSE dengan model yang digunakan. Sumbu Y mewakili nilai RMSE dan sumbu X mewakili model yang digunakan.



Gambar 3: RMSE model NB, GP, Poisson-Bayes dan CAR-BYM

Mangacu pada hasil perbandingan RMSE di atas, maka dapat diperoleh kesimpulan bahwa model CAR-BYM dapat memberikan hasil yang paling baik, yaitu

dengan menghasilkan nilai RMSE yang paling minimum diantara RMSE dari model-model yang dicobakan (Gambar 3).



Gambar 4: Grafik hubungan antara nilai RMSE dan nomor kecamatan untuk membandingkan nilai aktual dan nilai prediksi model NB, GP, Poisson-Bayes dan CAR-BYM

Selanjutnya pada model *negative binomial* (NB), model *generalized Poisson* (GP), dan model Poisson metode estimasi Bayes dapat dilihat bahwa model GP sedikit lebih baik. Gambar 4 merupakan perbandingan keakurasian hasil dari setiap model ketika masing-masing hasil prediksi dengan nilai aktualnya. Berdasarkan plot pada Gambar 4, dapat dilihat bahwa pola untuk nilai prediksi pada setiap model hampir mirip dan cenderung mengikuti pola data aktual. Namun demikian apabila dikaji lebih detail terkait dengan pola tersebut, maka dapat dilihat bahwa untuk pola yang berwarna hitam (CAR-BYM) dan warna biru (aktual) mempunyai kedekatan yang lebih dibandingkan dengan yang lainnya. Hal ini juga menguatkan dan sesuai dengan hasil RMSE sebagaimana disajikan pada Gambar 3.

4. Simpulan dan Saran

Metode bayes empiris dapat digunakan untuk mengestimasi proporsi persebaran TB di Kabupaten Banyumas. Sebagian besar wilayah kecamatan adalah dengan kasus 4-5 kasus per 1000 penduduk. Sementara untuk kasus TB yang berada di atas 7 per 10000 penduduk hanya satu kecamatan.

Pemodelan jumlah kasus tuberculosis (TB) di Kabupaten Banyumas mengalami overdispersi, sehingga model Poisson tidak relevan untuk digunakan. Model-model alternatif dalam penanganan masalah overdispersi yaitu NB, GP metode estimasi MLE dan model Poisson dan CAR-BYM metode estimasi Bayes. Model terbaik adalah model CAR-BYM dimana nilai RMSE yang paling kecil dibandingkan dengan model yang dibandingkan. Keakuratan model dengan nilai RMSE terkecil juga didukung dengan plot kesesuaian data aktual dan hasil prediksi paling mirip.

Penelitian berikutnya perlu dikembangkan matriks pembobot spasial yang diharapkan dapat meningkatkan akurasi ketepatan model. Beberapa konstruksi matriks pembobot spasial yang melibatkan ukuran autokorelasi spasial yang

mengakomodasi variabel yang diperhatikan dan kedekatan secara geografis perlu dipertimbangkan.

Ucapan Terima Kasih. Penulis mengucapkan terima kasih kepada Universitas Jenderal Soedirman melalui Hibah Peningkatan Kompetensi ini

Daftar Pustaka

- Aswi, A., Cramb, S., Duncan, E., & Mengersen, K. (2020). Evaluating the impact of a small number of areas on spatial estimation. *International Journal of Health Geographics*, 19(1): 1–14.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.
- Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4): 733–746.
- Cressie, N. A. (1993). *Statistics for spatial data*. John Willy and Sons. Inc., New York.
- De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, 64(1): 107–133.
- Gaetan, C., & Guyon, X. (2010). *Spatial statistics and modeling* (Vol. 90). Springer.
- Han, X., & Lee, L. (2013). Bayesian estimation and model selection for spatial Durbin error model with finite distributed lags. *Regional Science and Urban Economics*, 43(5): 816–837.
- Hendricks, J., & Neumann, C. (2020). A Bayesian approach for the analysis of error rate studies in forensic science. *Forensic Science International*, 306: 110047.
- Ibrahim, J., Chen, M., & Sinha, D. (2001). *Bayesian Survival Analysis* Springer Series in Statistics. New York, NY: Springer. Doi, 10: 978–1.
- Iddrisu, A.-K., & Amoako, Y. A. (2016). Spatial Modeling and Mapping of Tuberculosis Using Bayesian Hierarchical Approaches. *Open Journal of Statistics*, 6(3): 482–513.
- Kyung, M., & Ghosh, S. (2014). Maximum likelihood estimation for generalized conditionally autoregressive models of spatial data. *Journal of the Korean Statistical Society*, 43: 339–353.
- MacNab, Y. C., Farrell, P. J., Gustafson, P., & Wen, S. (2004). Estimation in Bayesian disease mapping. *Biometrics*, 60(4): 865–873.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (Vol. 37). CRC Press.
- Obaromi, D. (2019). Spatial Modelling of Some Conditional Autoregressive Priors in A Disease Mapping Model: the Bayesian Approach. *Biomedical Journal of Scientific & Technical Research*, 14(3).

- Srinivasan, R., & Venkatesan, P. (2014). Bayesian random effects model for disease mapping of relative risks. *Ann Biol Res*, 5(1): 23–31.
- Stern, H. S., & Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, 19(17-18): 2377–2397.