

Pemodelan Survival pada Kejadian Demam Berdarah Dengue di Surabaya Timur dengan Pendekatan Bayesian

Ni Luh Putu Ika Candrawengi, Nur Iriawan dan Irhamah
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: nur_i@statistika.its.ac.id

Abstrak—Salah satu penyakit yang paling sering dialami di Indonesia khususnya pada saat musim penghujan adalah penyakit Demam Berdarah Dengue (DBD). Penyakit ini disebabkan oleh virus *dengue* yang ditularkan oleh nyamuk *Aedes aegypti* betina. Untuk mengetahui laju kesembuhan pasien maka perlu diketahui faktor-faktor yang dapat mempengaruhinya. Salah satu metode yang dapat digunakan adalah analisis *survival*. Analisis *survival* merupakan analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu objek sampai terjadinya *event*. Dalam penelitian ini *event* yang digunakan adalah waktu hingga sembuh (lama rawat inap) pasien. Analisis Bayesian memperlakukan semua parameter yang tidak diketahui sebagai variabel *random* dan memiliki distribusi. Dalam mengestimasi parameter digunakan pendekatan bayesian untuk mengatasi kasus jumlah data terbatas karena pada bayesian menggunakan estimasi dengan mempertimbangkan distribusi *prior*. Digunakan dua macam distribusi yang berbeda untuk memodelkan *survival* untuk kasus demam berdarah *dengue*. Distribusi yang digunakan adalah distribusi *Generalized Gamma* dan *Weibull*. Model yang terbaik yang ditunjukkan dengan nilai WAIC terkecil adalah model *survival* dengan menggunakan distribusi *Weibull*.

Kata Kunci—Analisis Survival , Bayesian, Demam Berdarah Dengue, Kesembuhan.

I. PENDAHULUAN

INDONESIA merupakan negara iklim tropis yang memiliki dua musim yaitu musim penghujan dan musim kemarau. Musim penghujan dapat menimbulkan penyakit karena banyaknya virus, bakteri, jamur, dan parasit yang dapat berkembang akibat kelembaban yang tinggi. Salah satu virus yang cepat berkembang saat musim penghujan adalah virus *dengue*, apabila menjangkit manusia akan menimbulkan penyakit Demam Berdarah Dengue (DBD) yang kerap muncul di negara tropis seperti Indonesia [1].

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh virus *dengue* dengan gejala demam 2-7 hari disertai dengan nyeri otot dan atau nyeri sendi, leukopenia, ruam, lifadenopati, trombositopenia, dan diatesis hemoragik . Penyakit DBD ditularkan melalui gigitan nyamuk *Aedes aegypti*. Menurut data WHO, Indonesia dilaporkan sebagai negara ke-2 dengan kasus DBD terbesar diantara 30 negara wilayah endemis. Kasus Demam Berdarah Dengue (DBD) yang terjadi di Indonesia pada tahun 2017 mengalami penurunan signifikan dari tahun 2016 yang mulanya terdapat 204171 kasus menjadi 68407 kasus. Provinsi Jawa Timur merupakan salah satu provinsi dengan kasus DBD tertinggi

dengan jumlah kasus sebesar 7866 kasus. Pada tahun 2018 kasus DBD kembali meningkat dengan angka jumlah kasus sebanyak 9452 kasus. Kasus kematian akibat DBD yang terjadi di Indonesia pada tahun 2017 mengalami penurunan hampir tiga kali lipat. Pada tahun 2017 dan 2018 kasus kematian akibat DBD tertinggi terjadi di Provinsi Jawa Timur dengan banyak kematian 105 kematian [2]. Surabaya pada tahun 2018 mengalami penurunan kasus penyakit DBD. Walaupun terjadi penurunan kasus demam berdarah tetapi masih ada beberapa kecamatan di Surabaya yang memiliki angka kasus Demam Berdarah tinggi khususnya untuk wilayah Surabaya Timur.

Pada suatu data yang memiliki variabel respon berupa waktu terjadinya suatu *event* diperlukan suatu analisis yaitu analisis *survival*. Analisis *survival* adalah analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu objek sampai terjadinya *event*. Untuk mengetahui faktor-faktor yang berpengaruh terhadap laju kesembuhan pasien DBD, maka perlu dilakukan penelitian secara statistik untuk dapat mengurangi kasus kematian yang disebabkan oleh DBD. Hal ini dilakukan untuk mengetahui peluang pasien DBD untuk bertahan hidup, serta dalam metode penyembuhan beberapa pasien dapat bertahan dan seberapa besar kegagalan dari metode tersebut

Berdasarkan uraian tersebut, maka perlu dilakukan penelitian untuk mengkaji waktu kesembuhan pasien penyakit DBD berdasarkan faktor demografi, geografi, dan faktor *host* dari pasien DBD. Penelitian ini akan mengidentifikasi pola data pada waktu lama rawat inap untuk mengetahui model yang tepat untuk digunakan. Selanjutnya, dilakukan pemodelan *survival* berdasarkan faktor-faktor yang mempengaruhi laju kesembuhan pasien DBD yang diestimasi dengan Bayesian.

II. TINJAUAN PUSTAKA

A. Pengujian Kesesuaian Distribusi

Pemeriksaan kesesuaian antara pola distribusi data observasi dengan pola distribusi yang dipilih merupakan salah satu tujuan untuk melakukan pengujian kesesuaian distribusi. Beberapa metode yang sering digunakan untuk pengujian kesesuaian distribusi data, diantaranya yaitu *Anderson Darling*, *Kolmogorov-Smirnov*, dan *Chi-square*. Pada penelitian ini pengujian kesesuaian distribusi yang digunakan adalah menggunakan uji *Kolmogorov-Smirnov* (KS). Uji *Kolmogorov-Smirnov* (KS) merupakan pengujian yang didasarkan pada pendekatan distribusi kumulatif atau *Cumulative Density Function* (CDF). Selain itu, uji *Kolmogorov-Smirnov* (KS)

merupakan uji yang paling sering digunakan dalam pengujian distribusi[3]. Berikut merupakan Hipotesis uji *Kolmogorov Smirnov*:

$H_0 : F_n(x) = F_0(x)$ atau data mengikuti pola distribusi tertentu

$H_1 : F_n(x) \neq F_0(x)$ atau data tidak mengikuti pola distribusi tertentu

Nilai statistik uji *Kolmogorov-Smirnov* adalah sebagai berikut.

$$D = \max|F_n(x) - F_0(x)|, \quad (1)$$

dengan,

$F_0(x)$: fungsi distribusi yang dihipotesiskan,

$F_n(x)$: fungsi distribusi data secara empiris.

Daerah penolakan untuk hipotesis ini adalah jika nilai uji D lebih besar dari tabel *critical value Kolmogorov-Smirnov*. Selain menggunakan statistik uji seperti pada Persamaan (1), pengujian KS juga dapat dilihat dari *p-value*. Tolak H_0 apabila *p-value* kurang dari taraf signifikan α yang digunakan.

B. Analisis Survival

Analisis *survival* merupakan salah satu metode dalam statistika yang digunakan untuk menganalisis data yang memiliki variabel respon berupa waktu sampai terjadinya suatu *event*. *Event* merupakan suatu peristiwa ekstrim yang mungkin terjadi pada individu seperti kesembuhan, kekambuhan, ataupun kematian setelah dilakukan suatu proses *treatment*. Respon dari analisis *survival* didefinisikan sebagai *range* waktu dari awal penelitian hingga suatu *event* terjadi atau sampai penelitian berakhir [4]. Elemen-elemen yang harus diperhatikan dalam menentukan waktu *survival* adalah sebagai berikut :

1. *Time origin/starting point* atau merupakan titik awal, misalnya waktu awal dimulainya suatu *treatment*.
2. *Failure time* yaitu waktu berakhirnya *failure event*, misalnya tanggal keluar rumah sakit atau tanggal kematian
3. *Measurement scale of time* atau skala pengukuran waktu, misalnya hari atau bulan.

C. Fungsi Survival

Fungsi *survival* merupakan probabilitas bahwa suatu individu tidak mengalami kejadian lebih dari waktu t . Fungsi *survival* $S(t)$ dapat dinyatakan sebagai peluang seorang individu untuk bertahan lebih lama dari suatu waktu t [4]. Berikut merupakan persamaan untuk fungsi *survival*.

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad (2)$$

Fungsi *hazard* merupakan laju kegagalan (*failure*) sesaat ketika mengalami suatu *event* pada waktu ke- t . Berbeda dengan fungsi *survival* yang fokus pada keberhasilan, fungsi *hazard* fokus pada kegagalan ketika kejadian berlangsung. Fungsi *hazard* dapat dinyatakan dalam persamaan berikut.

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

Hubungan antara $H(t)$ atau fungsi kumulatif *hazard* dengan fungsi *survival* $S(t)$ adalah sebagai berikut :

$$H(t) = - \ln S(t) \quad (4)$$

dengan :

t : Waktu yang diamati,

T : Waktu *survival* seorang individu,

$F(t)$: Fungsi distribusi kumulatif (CDF) dari distribusi data,

$f(t)$: Fungsi kepadatan peluang (PDF),

$S(t)$: Fungsi *survival*,

$h(t)$: Fungsi *hazard* proporsional,

$H(t)$: Fungsi kumulatif *hazard*.

D. Asumsi Proporsional Hazard

Pengujian asumsi *proportional hazard* juga dapat dilakukan dengan menggunakan statistik uji *goodness of fit*. Adapun langkah-langkah untuk pengujian asumsi dengan uji *goodness of fit* adalah sebagai berikut.

1. Mendapatkan residual *schoenfeld* untuk setiap variabel prediktor dari model *cox*.
2. Membuat variabel *rank* waktu *survival* dari urutan kecil ke besar
3. Menghitung korelasi antara residual *schoenfeld* dan *rank* dari waktu *survival* menggunakan rumus pada Persamaan

$$r_{RT,PR_i} = \frac{\sum_{j=1}^n (PR_{ij} - P\bar{R})(RT_j - R\bar{T})}{\sqrt{\sum_{j=1}^n (PR_{ij} - P\bar{R})^2} \sqrt{\sum_{j=1}^n (RT_j - R\bar{T})^2}} \quad (5)$$

Keterangan :

PR_{ij} : residual *schoenfeld* untuk variabel prediktor ke- i individu yang mengalami *event* pada $t_{(j)}$.

RT_j : *rank* waktu *survival* yang diurutkan dari kecil ke besar pada $t_{(j)}$.

4. Melakukan pengujian korelasi antara residual *schoenfeld* dan *rank* waktu *survival* dengan hipotesis sebagai berikut.

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Statistik uji yang digunakan adalah sebagai berikut,

$$t_{hit} = \frac{r_{RT,PR_i} \sqrt{n-2}}{\sqrt{1-r_{RT,PR_i}^2}} \quad (6)$$

Tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2}, n-2}$ atau nilai *p-value* kurang dari taraf kesalahan (α) yang digunakan. Asumsi *proportional hazard* terpenuhi jika uji korelasi antara residual *schoenfeld* dan *rank* waktu *survival* tidak signifikan atau gagal tolak H_0 [4].

Perhitungan numerik dengan statistik uji global *Grambsch and Therneau* yang berdasarkan residual *schoenfeld* pada regresi *cox* dengan hipotesis sebagai berikut :

H_0 : Asumsi *proportional hazard* terpenuhi

H_1 : Asumsi *proportional hazard* tidak terpenuhi

Persamaan (7) merupakan statistik uji global variabel ke- p (G_p):

$$G_p = \frac{\sum_{i=1}^M \sum_{j=1}^{N_i} (\delta_{ij} h(t_{ij}) - \bar{h}) r_{s_{pij}}}{I_p \sum_{i=1}^M \sum_{j=1}^{N_i} (\delta_{ij} h(t_{ij}) - \bar{h})} \sim \chi^2_{(1)}, \quad (7)$$

dengan :

$$r_{spij} = \delta_{ij}\{x_{ij} - \bar{\alpha}_{pij}\}, \quad (8)$$

dimana,

$$\bar{\alpha}_{pij} = \frac{\sum_{I \in R_f} x_{pij} \exp(\beta x_i)}{\sum_{I \in R_f} \exp(\beta x_i)}, \quad (9)$$

δ_{ij} menunjukkan status kejadian (tersensor (0) dan tidak tersensor (1)), I_p merupakan elemen *information matrix* variabel ke- p , r_{spij} merupakan nilai residual *Schoenfeld* dan kesimpulan gagal tolak H_0 , jika nilai $G_p < \chi^2_{(1)}$ atau jika nilai $P\text{-value} > \alpha$ [5].

E. Model Proportional Hazard

Model *proportional hazard* atau *cox proportional hazard* (Cox PH) adalah suatu pemodelan matematis yang digunakan untuk menganalisis data *survival*. Pada model cox PH mampu mengestimasi koefisien regresi, *hazard ratio*, dan kurva *survival*. Cox PH merupakan model yang *robust* karena hasil dari model Cox PH mendekati aproksimasi hasil dari model parametrik [4]. Model cox PH merupakan model semi-parametrik karena model Cox PH tidak memerlukan informasi tentang distribusi yang mendasari waktu *survival* dan estimasi parameter regresi dari model cox PH tanpa harus menentukan baseline *hazard*. Adapun persamaan model Cox PH dapat ditulis seperti Persamaan 21[4].

$$h(t, X) = h_0(t) e^{\sum_{i=1}^p \beta_i x_i} \quad (10)$$

$$h(t, X) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (11)$$

dimana,

- (X_1, X_2, \dots, X_p) : Vektor yang berisi p variabel prediktor
- $h_0(t)$: *Baseline hazard* yang menggambarkan model *hazard* ketika semua variabel prediktornya bernilai nol
- $(\beta_1, \beta_2, \dots, \beta_p)$: Vektor pada parameter regresi

F. Analisis Bayesian

Estimasi parameter suatu model dapat digolongkan ke dalam dua kelompok yaitu data pengamatan saat ini yang bersifat sesaat selama studi dan data yang bersifat long memory histogram yang dikumpulkan di masa lampau [6]. Analisis Bayesian memperlakukan semua parameter yang tidak diketahui sebagai variabel *random* dan memiliki distribusi, sehingga analisis Bayesian berbeda dengan teori statistika klasik [7]. Dasar dari teorema Bayesian adalah distribusi posterior yang merupakan perkalian antara distribusi *prior* dengan data observasi penyusun fungsi *likelihood* [8].

Pada Penelitian ini, jika diketahui t adalah sebuah variabel *random* dengan fungsi densitas $f(t|\theta)$ dengan mengikuti pola distribusi tertentu. θ merupakan vektor parameter dengan $\theta = (\beta_1, \beta_2, \dots, \beta_p, \alpha, \gamma)$ yang memiliki ukuran sebanyak parameter yang digunakan, dalam penelitian ini parameter ukuran parameter yaitu $p + 1$, sedangkan t merupakan vektor sampel yang memiliki ukuran n dengan distribusi identik dan

independent. *Joint posterior* dari θ dan t dapat dituliskan dalam bentuk persamaan berikut.

$$f(\theta|t) = \frac{f(t|\theta)f(\theta)}{f(t)} \quad (12)$$

didapatkan *conditional distribution* yang dituliskan dalam persamaan berikut.

$$f(\theta|t) \propto f(t|\theta)f(\theta), \quad (13)$$

dengan :

$f(t|\theta)$: fungsi *likelihood*,

$f(\theta)$: *joint distribution* θ sebagai $f(\theta) =$

$f(\beta_1) \cdot f(\beta_2) \cdot \dots \cdot f(\beta_p) \cdot f(\alpha) \cdot f(\gamma)$,

$f(t)$: total probabilitas bayes yang berfungsi sebagai konstanta penormal fungsi distribusi densitas.

Penentuan distribusi *prior* sangat penting karena mempengaruhi bentuk distribusi posterior yang akan digunakan untuk pengambilan keputusan. Jika tidak terdapat informasi *prior* dapat digunakan *non-informative prior* dengan memberikan nilai yang tidak akan mempengaruhi distribusi posterior. Berikut ini merupakan macam-macam distribusi *prior*, antara lain :

- a. *Informative* dan *noninformative prior* merupakan penentuan *prior* yang didasarkan pada ada atau tidaknya informasi yang diperoleh dari penelitian sebelumnya mengenai pola distribusi data.
- b. *Conjugate* dan *nonconjugate prior* merupakan penentuan *prior* yang didasarkan dari pola *likelihood* data.
- c. *Proper* dan *improper prior* merupakan penentuan *prior* tergantung pada pemberian pembobotan/densitas di setiap titik, apakah terdistribusi secara uniform atau tidak
- d. *Pseudo prior* merupakan penentuan *prior* dengan nilai yang pemberian nilainya didasarkan dan disetarakan dengan hasil elaborasi cara *frequentist*.

Distribusi posterior merupakan distribusi peluang yang berisi informasi mengenai parameter θ atau dapat dinyatakan dalam $f(\theta|t)$. Distribusi *posterior* merupakan gabungan antara informasi *prior* dan data observasi (fungsi *likelihood*). *Markov Chain Monte Carlo* (MCMC) merupakan suatu pendekatan yang dapat digunakan untuk memecahkan masalah dalam hal pembaharuan parameter dari sampel melalui distribusi posterior. *Markov Chain Monte Carlo* (MCMC) dapat digunakan untuk membentuk model yang sangat kompleks, berdimensi tinggi atau memiliki sifat data yang berkorelasi tinggi. Untuk mendapatkan distribusi *prior* yang stasioner, MCMC memiliki ide dasar dengan membangkitkan sebuah *Markov Chain* dengan simulasi *Monte Carlo* yang beriterasi. Salah satu algoritma dari MCMC yang mencakup iteratif sampling dari setiap distribusi bersyarat adalah *Gibbs*. Algoritma *Gibbs sampling* dapat digunakan walaupun data memiliki distribusi yang rumit dan tidak diberitahu secara eksplisit asalkan distribusi kondisional bersyarat univariat [9]. Diberikan $\theta = (\beta_1, \beta_2, \dots, \beta_p, \alpha, \gamma)$, dimana bentuk distribusi *full conditional* yang digunakan untuk masing-masing parameter adalah: *sampling*. Algoritma *gibbs sampling* digunakan untuk mengambil sampel dari distribusi kompleks berdimensi tinggi.

$$p(\alpha|x, \beta_2, \dots, \beta_p, \gamma), p(\gamma|x, \beta_2, \dots, \beta_p, \alpha), \\ p(\beta_1|x, \beta_2, \dots, \beta_p, \alpha, \gamma), \dots, p(\beta_p|x, \beta_1, \dots, \beta_{p-1}, \alpha, \gamma))$$

atau dapat ditulis dengan $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(p)}$. Berikut ini merupakan langkah-langkah dari algoritma gibbs sampling :

1. Mengambil nilai $m = 0$ dan menentukan nilai inisiasi (*initial value*) dari $\theta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)}, \alpha^{(0)}, \gamma^{(0)})$.
2. Membangkitkan tiap komponen dari $\theta^{(m+1)} = (\beta_1^{(m+1)}, \beta_2^{(m+1)}, \dots, \beta_p^{(m+1)}, \alpha^{(m+1)}, \gamma^{(m+1)})$ dimana,
 Nilai $\alpha^{(m+1)}$ berasal dari $p(\beta_1 | \mathbf{X}, \beta_1^{(m)}, \dots, \beta_p^{(m)}, \gamma^{(m)})$
 Nilai $\beta_1^{(m+1)}$ berasal dari $p(\beta_1 | \mathbf{X}, \beta_1^{(m)}, \dots, \beta_p^{(m)}, \alpha^{(m+1)})$
 Nilai $\beta_2^{(m+1)}$ berasal dari $p(\beta_2 | \mathbf{X}, \beta_2^{(m)}, \dots, \beta_p^{(m)}, \alpha^{(m+1)}, \gamma^{(m+1)})$
 Nilai $\beta_p^{(m+1)}$ berasal dari $p(\beta_p | \mathbf{X}, \beta_1^{(m+1)}, \dots, \beta_{p-1}^{(m+1)}, \alpha^{(m+1)}, \gamma^{(m+1)})$
3. Melakukan monitoring terhadap kekonvergenan algoritma. Apabila belum konvergen, maka dibangkitkan lebih banyak observasi dengan mengambil nilai $m_1 = m + 1, m_2 = m + 1, \dots, m_k = m + ka$ dan diulangi Langkah 1 dan Langkah 2.
4. Membuang b observasi pertama (*burn in period*). Hal ini dilakukan untuk menghindari pengaruh *initial value*.
5. Anggap $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(p)}$ dan sebanyak b observasi pertama sudah dibuang dan diperoleh sampel $\theta^{(b+1)}, \theta^{(b+2)}, \dots, \theta^{(k)}$ untuk analisis posterior,.
6. Membuat plot distribusi posterior.
7. Mendapatkan mean, median, standar deviasi, kuantil, dan MC error dari posterior.

G. Watanabe Akaike Information Criterion (WAIC)

Watanabe Akaike Information Criterion (WAIC) atau yang biasa dikenal dengan *Widely Applicable Information Criterion* adalah salah satu pendekatan bayesian untuk mengestimasi ekspektasi *out-of-sample*. Metode ini dimulai dengan menghitung log dari fungsi densitas posterior dan menambahkan koreksi untuk jumlah parameter efektif yang disesuaikan untuk *overfitting*. Dibandingkan dengan DIC dan AIC, WAIC memiliki sifat yang berada diatas rata-rata diatas distribusi posterior dibandingkan dalam satu estimasi titik. Hal ini sangat relevan digunakan dalam prediksi karena WAIC mengevaluasi prediksi yang sebenarnya digunakan untuk data baru dalam konteks Bayesian [10]. Rumus untuk menghitung WAIC dapat dilihat pada Persamaan (14).

$$WAIC = lppd - pWAIC, \tag{14}$$

dengan,

$$pWAIC = \sum_{i=1}^n var_{post} \log p(y_i | \theta) \tag{15}$$

H. Demam Berdarah Dengue (DBD)

Demam Berdarah *Dengue* (DBD) adalah penyakit menular yang disebabkan oleh virus *dengue* dan ditularkan oleh nyamuk *Aedes aegypti*. Biasanya, penyakit demam berdarah ini ditandai dengan adanya demam mendadak selama 2-7 hari, lemah/lesu, nyeri ulu hati, nyeri persendian, dan dapat disertai dengan tanda perdarahan dikulit berupa bintik-bintik. Jika penyakit

Tabel 1.
Variabel Penelitian

Variabel	Nama Variabel	Skala
t	Lama rawat inap pasien	Rasio
t.cen	data tersensor	Nominal
X ₁	Jenis Kelamin Pasien 0 : Perempuan 1 : Laki-laki	Nominal
X ₂	Usia pasien saat dirawat	Rasio
X ₃	Pendidikan terakhir pasien 1: SD 2 : SMP 3: SMA 4 : S1/D3 5 : Lain-lain	Ordinal
X ₄	Pekerjaan pasien : 1: Ibu Rumah Tangga 2: Pegawai Swasta 3: Pegawai Negeri 4: Tidak Bekerja 5: Lain-lain	Nominal
X ₅	Tingkatan/fase DBD yang diidap : 1: I 2: II 3: III 1. IV	Ordinal
X ₆	Lama demam sebelum masuk ke RS	Rasio
X ₇	Banyaknya denyut nadi per menit	Rasio
X ₈	Frekuensi pernafasan pasien	Rasio
X ₉	Suhu badan pasien saat pertama kali datang ke rumah sakit	Rasio
X ₁₀	Kadar Hemoglobin Pasien	Rasio
X ₁₁	Kadar Hematokrit Pasien	Rasio
X ₁₂	Kadar sel darah putih pasien	Rasio
X ₁₃	Kadar trombosit dalam darah	Rasio

demam berdarah sudah sangat parah maka dapat disertai dengan mimisan, muntah darah, *shock*, bahkan dapat menimbulkan kematian. Tiga faktor penting yang memegang peranan pada penularan infeksi virus *dengue* yaitu manusia, virus, dan vektor perantara. Virus *dengue* ditularkan kepada manusia melalui gigitan nyamuk *Aedes aegypti* betina. Tidak semua orang yang didalam tubuhnya terdapat virus *dengue* akan sakit demam berdarah *dengue*. Beberapa akan mengalami demam ringan dan sembuh dengan sendirinya atau bahkan ada yang sama sekali tanpa gejala sakit.

III. METODOLOGI PENELITIAN

A. Sumber Data

Pada penelitian ini sumber data yang akan digunakan merupakan data sekunder yang diperoleh dari data rekam medis pasien rawat inap DBD di RSUD Dr. Soetomo Surabaya, waktu rekam medis dimulai dari tahun 2016 sampai 2019. Waktu penelitian adalah 1 Januari 2016 sampai dengan 31 Desember 2019.

B. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini terdapat pada Tabel 1.

Tabel 2.
Hasil Uji Proporsional Hazard

Variabel	rho	P-value
Jenis Kelamin (1)	0,05481	0,8374
Usia Pasien	0,41291	0,1397
Pendidikan Terakhir (S1/D3)	-0,2339-	0,2671
Pendidikan (SMA)	Terakhir -0,20646	0,3335
Pekerjaan Negeri	Pegawai 0,09771	0,6676
Pekerjaan Swasta	(Pegawai -0,11479)	0,6045
Pekerjaan Bekerja	(Tidak -0,32063)	0,1018
Lama demam masuk RS	sebelum 0,33498	0,0936
Fase DBD (II)	-0,20733	0,3648
Fase DBD (III)	-0,01608	0,9337
Denyut Nadi	-0,24943	0,1697
Pernafasan	0,10119	0,6304
Suhu	-0,13837	0,5282
HGB	-0,00966	0,9645
HCT	0,16331	0,3972
WBC	-0,24273	0,2864
PLT	-0,31504	0,0627
GLOBAL	NA	0,8076

Tabel 3.

Distribusi	Uji Distribusi Waktu <i>Survival</i>		
	Statistik Uji	Nilai Kritis	Keputusan
<i>Generalized</i>	0,13276	0,28724	Gagal tolak H0
<i>Weibull</i>	0,15079	0,28724	Gagal tolak H0

C. Langkah Analisis

Langkah-langkah analisis dalam penelitian ini adalah :

- Melakukan pengujian distribusi waktu rawat inap.
- Melakukan pengujian asumsi pemodelan *proporsional hazard*.
- Proses estimasi parameter untuk distribusi dari variabel respon lama rawat inap dengan menggunakan MCMC dan *gibbs sampler*.
- Menentukan model *survival* spasial pada kejadian DBD di Surabaya Timur berdasarkan faktor-faktor yang mempengaruhi laju kesembuhan pasien menggunakan program OpenBUGS. Langkah-langkahnya dapat dijelaskan sebagai berikut.
 - Menentukan distribusi *prior* yang digunakan berdasarkan informasi dari data
 - Menentukan *joint posterior* model *survival*
 - Menentukan *full conditional* untuk setiap parameter
 - Melakukan inisiasi untuk parameter awal
 - Menghitung nilai fungsi *hazard* dan *survival* pasien DBD berdasarkan posterior *summaries* yang diperoleh
- Interpretasi model *survival* sehingga diperoleh faktor-faktor yang berpengaruh terhadap laju kesembuhan pasien DBD.
- Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

A. Pendugaan Distribusi Lama Rawat Inap (Waktu Survival) Pasien DBD

Pendugaan distribusi dilakukan terhadap waktu *survival* (*t*) yaitu lama rawat inap pasien DBD Surabaya Timur di RSUD Dr. Soetomo. Oleh sebab itu, dalam pengujian kesesuaian distribusi digunakan asumsi awal yaitu data dari lama rawat

inap pasien DBD berdistribusi *Generalized* Gamma 3P/Weibul 2P.

Hasil uji kesesuaian distribusi pada Tabel 2 menunjukkan bahwa waktu lama rawat inap (waktu *survival*) berdistribusi *Generalized* Gamma dan Weibull. Hal ini disebabkan karena nilai statistik uji *Kolmogorov-Smirnov* dari data lama rawat inap pasien DBD lebih kecil dibandingkan dengan nilai kritis yang ditetapkan sebesar 0,28274. Selain itu, lama waktu rawat inap pasien (waktu *survival*) selalu memiliki nilai yang cenderung positif sehingga distribusi yang sesuai yaitu *Generalized* Gamma dan distribusi Weibull.

B. Asumsi *Proportional Hazard*

Uji global *Grambsch and Thernau* juga digunakan dalam penelitian ini. Pengujian ini menggunakan hipotesis serta statistik uji sesuai dengan Persamaan (5). Tabel 3 Merupakan hasil uji *Schoenfeld residual*.

Hasil pengujian *proporsional hazard* dengan menggunakan korelasi antara *residual Schoenfeld* dengan waktu *survival* yang ditunjukkan dengan rho menunjukkan untuk seluruh variabel kategorik yang digunakan dalam penelitian ini memiliki nilai *P* yang lebih dari taraf signifikansi sebesar 0,05. Sehingga, untuk seluruh variabel kategorik tidak signifikan sehingga gagal menolak hipotesis untuk asumsi *proporsional hazard*. Hasil Uji Global untuk menguji asumsi *proporsional hazard* untuk keseluruhan model secara simultan menunjukkan nilai *P* sebesar 0,8076 yang lebih dari taraf signifikansi. Dari hasil uji Global dapat disimpulkan bahwa asumsi *proporsional hazard* terpenuhi pada model ini

C. Penentuan *Prior*

Pada penelitian ini digunakan dua macam distribusi untuk lama inap (waktu *survival*) pasien DBD yaitu distribusi Weibull 2 Parameter dan *Generalized* Gamma 3 parameter. Fungsi kepadatan peluang (PDF) untuk distribusi Weibull 2 parameter adalah sebagai berikut,

$$f(t|\gamma, \mu) = \mu \gamma t^{\gamma-1} \exp(-\mu t^\gamma).$$

Kemudian fungsi *hazard* untuk distribusi Weibull adalah sebagai berikut,

$$h(t) = \gamma t^{\gamma-1} \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p),$$

sebelum melakukan estimasi terhadap parameter γ dan untuk β , maka terlebih dahulu menentukan distribusi *prior*nya sebagai berikut,

$$\gamma \sim \text{Gamma}(a, b),$$

$$\beta_i \sim \text{Normal}(m, \text{tau}).$$

Untuk distribusi *Generalized* Gamma diketahui fungsi kepadatan peluangnya adalah sebagai berikut,

$$f(t|\gamma, \alpha, \beta) = \frac{\beta}{\Gamma(\alpha)} \frac{t^{\beta\alpha-1}}{\gamma^{\beta\alpha}} \exp\left(-\left(\frac{t}{\gamma}\right)^\beta\right),$$

Tabel 4.
Nilai WAIC Masing-masing Model *Survival*

Model	WAIC
Model <i>Survival</i> distribusi <i>Generalized Gamma</i>	147,1
Model <i>survival</i> distribusi Weibull	93,01

dengan model *hazard*,

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p).$$

Dimana $h_0(t)$ dari distribusi *Generalized Gamma* merupakan bentuk integral rumit yang memuat parameter α, θ dan γ . Sebelum melakukan estimasi terhadap parameter α, θ dan untuk β , maka terlebih dahulu menentukan distribusi *prior*nya sebagai berikut,

$$\begin{aligned} \alpha &\sim \text{Gamma}(a, b), \\ \gamma &\sim \text{Gamma}(a, b), \\ \beta_i &\sim \text{Normal}(m, \tau), \end{aligned}$$

dimana penentuan distribusi *prior* tersebut dilakukan berdasarkan gabungan antara *prior* informatif dan *pseudo prior*.

D. Faktor-faktor yang Signifikan Mempengaruhi Laju Kesembuhan Pasien DBD

1) Faktor yang Signifikan Mempengaruhi Laju Kesembuhan Pasien DBD dengan Distribusi Weibull

Faktor-faktor yang dianggap signifikan untuk mempengaruhi laju kesembuhan pasien DBD di RSUD Dr. Soetomo adalah jika nilai dari *credible interval* 2,5% hingga 97,5% tidak memuat nilai 0. Tidak semua variabel mempengaruhi laju kesembuhan pasien DBD di RSUD Dr. Soetomo. Variabel yang berpengaruh terhadap kesembuhan pasien DBD adalah variabel jenis kelamin, pendidikan terakhir pegawai negeri dan lain-lain, diagnosis DB ketika masuk ke rumah sakit, lama demam sebelum masuk rumah sakit, denyut nadi per menit, tingkat pernafasan, kadar *White Blood Cells*, dan kadar *platelets* atau trombosit.

2) Faktor yang Signifikan Mempengaruhi Laju Kesembuhan Pasien DBD dengan Distribusi *Generalized Gamma*

Faktor-faktor yang dianggap signifikan untuk mempengaruhi laju kesembuhan pasien DBD di RSUD Dr. Soetomo jika diduga bahwa lama rawat inap pasien mengikuti pola distribusi *Generalized Gamma* adalah variabel jenis kelamin, pendidikan terakhir untuk kategori S1, pekerjaan untuk kategori pegawai negeri, pegawai swasta, dan lain-lain, fase DBD, lama demam sebelum masuk rumah sakit, fase DBD yang didiagnosis II, denyut nadi per menit, tingkat pernafasan, kadar *white blood cells*, dan kadar *platelets* atau trombosit.

E. Laju Kesembuhan Pasien DBD

Hazard rate (laju kesembuhan) pasien DBD RSUD Dr. Soetomo di masing-masing kecamatan yang berada di daerah Surabaya Timur dapat dimodelkan sebagai berikut. Model Laju Kesembuhan Pasien Distribusi Weibull

$$h(t) = 2,368t^{1,368} \exp(-3,898 + 0,6833X_{1(perempuan)} + 0,3896X_{3(S1/D3)})$$

$$\begin{aligned} &-0,5509X_{4(P.Negeri)} + 0,4727X_{4(lain-lain)} - 0,4902X_{5(I)} \\ &\quad - 0,462X_6 \\ &+ 0,74X_7 - 0,5048X_8 - 0,6802X_{12} + 0,7391X_{13} \end{aligned}$$

Model Laju Kesembuhan Pasien Distribusi *Generalized Gamma*

$$\begin{aligned} h(t) = h_0(t) \exp(&-2,154 + 0,7327X_{1(perempuan)} \\ &+ 0,3812X_{3(S1/D3)} - 0,4973X_{4(P.Negeri)} \\ &- 0,5489X_{4(P.Swasta)} + 0,5492X_{4(lain-lain)} - 0,4938X_{5(I)} \\ &\quad - 0,49X_6 + 0,7913X_7 \\ &- 0,5521X_8 - 0,626X_{12} + 0,6793X_{13}) \end{aligned}$$

F. Perbandingan Model *Survival*

WAIC merupakan bentuk *Generalized* nilai *Akaike Information Criterion* (AIC) ke dalam bentuk model statistika yang singular. Model terbaik yang dipilih adalah model yang memiliki nilai WAIC yang lebih kecil. Berikut ini pada Tabel 4 merupakan nilai WAIC dari masing-masing model *survival*. Berdasarkan Tabel 4 maka model yang dipilih adalah model *survival* dengan distribusi Weibull. Pemilihan ini dikarenakan nilai WAIC dari model tersebut sebesar 93,01 yang merupakan nilai WAIC terkecil dari model yang ada. Oleh sebab itu maka laju kesembuhan pasien DBD di Surabaya Timur dapat dimodelkan dengan model sebagai berikut.

$$\begin{aligned} \lambda(t) = 2,368t^{1,368} \exp(&-3,898 \\ &+ 0,6833X_{1(perempuan)} \\ &+ 0,3896X_{3(S1/D3)} \\ &- 0,5509X_{4(P.Negeri)} + 0,4727X_{4(lain-lain)} \\ &\quad - 0,4902X_{5(I)} - 0,462X_6 \\ &+ 0,74X_7 - 0,5048X_8 - 0,6802X_{12} + 0,7391X_{13}) \end{aligned}$$

V. KESIMPULAN DAN SARAN

Kesimpulan yang diperoleh dari hasil analisis yang telah dilakukan adalah sebagai berikut; (1) Distribusi dari lama rawat inap pasien DBD mengikuti pola distribusi *Generalized Gamma* atau Weibull; (2) Pada penelitian model yang dipilih untuk memodelkan laju kesembuhan pasien DBD adalah model *survival* dengan distribusi Weibull. Faktor-faktor yang mempengaruhi laju kesembuhan pasien adalah variabel jenis kelamin, pendidikan terakhir pasien untuk kategori S1/D3, pekerjaan pasien untuk kategori pegawai negeri dan lain-lain, diagnosis DB ketika masuk ke rumah sakit, lama demam sebelum masuk rumah sakit, denyut nadi per menit, tingkat pernafasan, kadar *White Blood Cells*, dan kadar *platelets* atau trombosit.

Berdasarkan kesimpulan yang diperoleh, dapat dirumuskan saran sebagai berikut. (1) Untuk memodelkan dapat dipertimbangkan untuk kasus *mixture* pada data sehingga dapat menghasilkan hasil yang lebih baik; (2) Dapat dipertimbangkan untuk menambahkan efek spasial untuk penelitian selanjutnya; (3) Untuk pemerintah Kota Surabaya dapat mempertimbangkan untuk pencegahan penyakit DBD di beberapa kecamatan yang ada di Kota Surabaya.

DAFTAR PUSTAKA

- [1] W. H. O. WHO, "Dengue increase likely during rainy season: WHO warns," *World Health Organization (WHO)*, 2019. <https://www.who.int/westernpacific/news/detail/11-06-2019-dengue-increase-likely-during-rainy-season-who-warns>.
- [2] K. R. I. Ditjen P2P, "Situasi Penyakit Demam Berdarah di Indonesia Tahun 2017," 2018.
- [3] S. Siegel and J. J. N. Castellan, "Nonparametric Statistics for the Behavioral Sciences," *New York, Springer-Verlag*, 1988.
- [4] D. G. Kleinbaum and M. Klein, "Survival Analysis: A self-learning text, 2005," *New York, Springer-Verlag*, 2011.
- [5] P. M. Grambsch and T. M. Therneau, "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, vol. 81, no. 3, pp. 515–526, 1994, doi: 10.1093/biomet/81.3.515.
- [6] N. Iriawan, "Studi tentang 'Bayesian Mixture Normal'dengan menggunakan metode Markov Chain Monte Carlo (MCMC)," *Lap. Penelit.*, 2001.
- [7] W. M. Boldstat, *Introduction to Bayesian Statistics*, 2nd ed. New Jersey: John Wiley & Sons. Inc, 2007.
- [8] G. E. P. Box and G. C. Tiao, "Bayesian Inference in Statistical Analysis, Addison-Wesley," *Reading, MA*, 1973.
- [9] W. R. Gilks and P. Wild, "Adaptive Rejection Sampling for Gibbs Sampling," *Appl. Stat.*, vol. 41, no. 2, p. 337, 1992, doi: 10.2307/2347565.
- [10] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for Bayesian models," *Stat. Comput.*, vol. 24, no. 6, pp. 997–1016, 2014.