

Pemodelan Kasus Covid-19 di Jawa Timur Menggunakan Metode *Generalized Poisson Regression* dan *Negative Binomial Regression*

Syaillendra Ardifasalma, dan Ulil Azmi

Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: ulil_azmi@statistika.its.ac.id

Abstrak—Virus SARS-CoV-2 atau juga dikenal sebagai COVID-19, pertama kali ditemukan di China pada akhir 2019 dan telah menyebar secara global dan menyebabkan lebih dari 178 juta kasus terkonfirmasi dan sebanyak 3,9 juta jiwa meninggal dunia. Untuk kasus di Jawa Timur sendiri kasus COVID-19, hingga bulan Januari 2022 jumlah kasus di Jawa Timur yang terpapar virus COVID-19 sendiri mencapai 402.879 jiwa, sedangkan jumlah yang sembuh mencapai 371.745 jiwa dan meninggal dunia sebanyak 29.774 jiwa. Analisis regresi menggunakan variabel dependen sebagai variabel acak kontinu untuk menganalisis data. Sedangkan Regresi Poisson sendiri merupakan model dengan variabel Y berdistribusi Poisson. Namun dalam model regresi Poisson asumsi sering dilanggar antara estimasi varians yang berada di atas mean (*over-dispersion*) atau di bawah mean (*under-dispersion*). Salah satu model yang digunakan untuk menangani *under-dispersi* atau *over-dispersi* ini yaitu *Generalized Poisson Regression* dan *Negative Binomial Regression*. Data yang akan digunakan untuk meramalkan jumlah korban jiwa COVID-19 yaitu data harian di Jawa Timur dari bulan Oktober 2020 sampai dengan Januari 2022. Proses analisis data dilakukan dengan menggunakan *software RStudio* dengan faktor yang diduga mempengaruhi yaitu, kasus aktif, kasus baru, *Stringency Index*, dan *Bed Occupancy Rate* di Provinsi Jawa Timur. Penelitian ini diharapkan dapat membantu Satuan Tugas COVID-19 Provinsi Jawa Timur dalam pengambilan kebijakan untuk mengantisipasi jumlah pasien berdasarkan faktor – faktor yang berpengaruh signifikan pada penelitian dan juga menambah wawasan mengenai faktor –faktor apa saja yang dapat berpengaruh dengan kasus COVID-19 sehingga masyarakat bisa lebih waspada lagi dalam masa pandemi ini. Hasil penelitian menunjukkan bahwa model terbaik adalah model *Generalized Poisson Regression*. Hal ini ditunjukkan dari nilai $-2\log L$, AIC, dan BIC pada model *Generalized Poisson Regression* yang lebih kecil daripada model regresi poisson dan model *Negative Binomial Regression*. Dengan faktor-faktor yang berpengaruh terhadap jumlah kasus meninggal COVID-19 adalah kasus aktif, kasus baru, *Stringency Index*, dan *Bed Occupancy Rate*.

Kata Kunci—COVID-19, *Generalized Poisson Regression*, *Negative Binomial Regression*, AIC.

I. PENDAHULUAN

S AAT ini kita sedang mengalami pandemi, yang disebabkan oleh mewabahnya virus SARS-CoV-2 atau juga dikenal sebagai COVID-19, pertama kali ditemukan di China pada akhir 2019 dan telah menyebar secara global dan menyebabkan lebih dari 178 juta kasus terkonfirmasi dan sebanyak 3,9 juta jiwa meninggal dunia. Wuhan merupakan tempat di mana infeksi COVID-19 [1]. Pada akhir Mei 2020, Pusat Pengendalian dan Pencegahan Penyakit AS (CDC) memperkirakan bahwa sekitar sepertiga dari infeksi virus corona tidak menunjukkan gejala dan sekitar 0,4% orang yang menunjukkan gejala meninggal, menyiratkan tingkat

kematian infeksi (IFR) sebesar 0,3% (CDC 2020) [2]. Untuk kasus di Jawa Timur sendiri kasus COVID-19 pertama kali ditemukan pada tanggal 19 Oktober 2020 sebanyak 9 jiwa, hingga bulan Januari 2022 jumlah kasus di Jawa Timur yang terpapar virus COVID-19 sendiri mencapai 402.879 jiwa, sedangkan jumlah yang sembuh mencapai 371.745 jiwa dan meninggal dunia sebanyak 29.774 jiwa.

Model regresi Poisson telah banyak digunakan untuk menganalisis menghitung data di mana rata-rata sampel dan varians sampel hampir sama. Juga diketahui bahwa hitungan sering kali menunjukkan varians sampel lebih besar atau lebih kecil dari rata-rata sampel dan diklasifikasikan sebagai *overdispersi* atau *underdispersi* [3]. Peneliti tertarik untuk menerapkan metode *Generalized Poisson Regression* dan *Negative Binomial Regression* pada pemodelan kasus COVID-19 di Provinsi Jawa Timur dengan menggunakan data harian di Jawa Timur dari bulan Oktober 2020 sampai dengan Januari 2022. Faktor yang diduga mempengaruhi jumlah korban COVID-19 yaitu kasus aktif COVID-19, kasus baru COVID-19, indeks keketatan respons pemerintah, dan *Bed Occupancy Rate* di Provinsi Jawa Timur.

II. TINJAUAN PUSTAKA

A. *Coronavirus Disease 2019* (COVID-19)

Coronavirus Disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh *Coronavirus* jenis baru. Berdasarkan hasil penyelidikan epidemiologi, kasus yang terjadi diduga berhubungan dengan Pasar Seafood yang berada di Wuhan, China. Pada 7 Januari 2020, Pemerintah China mengumumkan bahwa penyakit tersebut disebabkan oleh *Coronavirus* jenis baru yang selanjutnya diberi nama SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*). Virus ini berasal dari famili yang sama dengan virus penyebab SARS dan MERS, tetapi SARS-CoV-2 lebih menular. Oleh karena itu, diperlukan intervensi tidak hanya dari sisi penerapan protokol kesehatan, tetapi juga intervensi yang efektif untuk memutus rantai penularan penyakit COVID-19. [4].

B. *Keluarga Exponensial*

Keluarga eksponensial adalah keluarga distribusi yang sering digunakan pada statistika parametrik. Distribusi yang masuk dalam keluarga eksponensial terdiri dari distribusi diskrit maupun distribusi kontinu, contohnya distribusi normal, Bernoulli, Binomial, Multinomial, Eksponensial, dan Poisson. Sebuah keluarga parametrik dari distribusi kontinu univariat dikatakan sebagai keluarga eksponensial jika dan hanya jika fungsi kepadatan probabilitas dari setiap anggota keluarga dapat ditulis seperti pada persamaan (1).

Tabel 5.
Statistik deskripsi kasus COVID-19

Variabel	Observasi	Rata-rata	Std. Dev.	Min.	Max.
Meninggal	488	54.47	84.801	0	431
Kasus Aktif	488	6618.08	11801.784	91	57126
Kasus Baru	488	735.93	1264.638	5	8230
Stringency Index	488	66.5437	6.368	50.46	72.69
BOR	488	41.9905	29.10754	1.40	91.60

Tabel 6.
Nilai VIF dari variabel prediktor

Variabel Prediktor	VIF
Aktif	5.695
Kasus Baru	6.039
Stringency Index	1.125
BOR	1.198

Tabel 7.
Estimasi parameter pada regresi poisson

Parameter	Koefisien	Std. error	Z	$P > z $
Meninggal	-1.657	0.1227	-13.506	0.00
Kasus Aktif	4.774×10^{-5}	6.249×10^{-7}	76.404	0.00
Kasus Baru	-2.907×10^{-5}	6.137×10^{-6}	-4.737	0.00
Stringency Index	0.067	1.798×10^{-5}	37.211	0.00
BOR	0.009	2.93×10^{-4}	31.486	0.00

$$f(x) = h(x) \exp[\theta^T t(x) - a(\theta)] \quad (1)$$

C. Maximum Likelihood Estimation

Dalam menentukan penaksir maksimum *likelihood*, membutuhkan fungsi *likelihood* dan kemudian memaksimumkan fungsi *likelihood* terhadap parameter yang akan ditaksir. Dalam beberapa kasus, kadang kala dikerjakan dengan melakukan transformasi logaritma natural (ln) terhadap fungsi *likelihood*, selanjutnya disebut sebagai log fungsi *likelihood*. Prosedur menentukan penaksir maksimum *likelihood* adalah sebagai berikut:

1. Menentukan fungsi *likelihood*, $L(\theta)$.
2. Melakukan transformasi logaritma natural dari $L(\theta)$.
3. Melakukan diferensial $\ln L(\theta)$ terhadap θ , dan kemudian menyamakan hasil dengan nol (persamaan *likelihood*).

Menentukan akar persamaan *likelihood* (θ), yang disebut sebagai penaksir maksimum *likelihood* (θ).

D. Regresi Poisson

Regresi Poisson adalah analisis regresi nonlinier dari distribusi Poisson yang digunakan untuk memprediksi ketergantungan variabel yang terdiri dari data diskrit atau jumlah yang diberikan satu atau lebih variabel bebas. Variabel yang diprediksi adalah disebut variabel dependen dan variabel yang digunakan untuk memprediksi nilai variabel dependen disebut variabel independen [5-6]. Salah satu distribusi yang berasal dari keluarga distribusi eksponensial merupakan distribusi Poisson[7]. Dimana Y menjadi variabel acak (laju di mana kasus COVID-19 terjadi) dan hasil dari kasus menjadi sebuah peristiwa. Variabel Y dikatakan mengikuti Distribusi Poisson dengan parameter $\lambda > 0$ jika memiliki fungsi probabilitas [8]. Sehingga didapatkan.

Tabel 1.
Nilai devians dan pearson model regresi poisson

Kriteria	Nilai	d.f.	Nilai/df
Deviance	7871.759	483	16.298
Pearson Chi-Square	8393.767	483	17.378

Tabel 2.
Estimasi parameter pada *generalized poisson regression*

Parameter	Koefisien	Std. error	Z	$P > z $
Meninggal	-1.559	0.3498	-4.457	0.00
Kasus Aktif	4.957×10^{-5}	2.76×10^{-6}	17.962	0.00
Kasus Baru	-7.214×10^{-5}	2.686×10^{-5}	-2.686	0.00
Stringency Index	0.0369	5.223×10^{-3}	7.079	0.00
BOR	0.01848	1.083×10^{-3}	17.058	0.00

Tabel 3.
Estimasi parameter pada *negative binomia regression*

Parameter	Koefisien	Std. error	Z	$P > z $
Meninggal	-1.557	0.349	-4.457	0.00
Kasus Aktif	8.661×10^{-5}	5.968×10^{-6}	14.512	0.00
Kasus Baru	-1.899×10^{-4}	5.737×10^{-5}	-3.31	0.00
Stringency Index	5.382×10^{-2}	5.294×10^{-3}	10.166	0.00
BOR	2.003×10^{-2}	1.191×10^{-3}	16.821	0.00

Tabel 4.
Pemilihan Model Terbaik Berdasarkan Model kasus COVID-19

Models	AIC	BIC
PR	10217.412	10238.363
NBR	4103.979	4129.121
GPR	4061.512	4086.654

$$P(y, \lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \quad (2)$$

E. Multikolinieritas

Multikolinieritas merupakan situasi yang menunjukkan adanya korelasi yang tinggi di antara variabel yang digunakan dalam model. Dalam model regresi, adanya korelasi antar variabel prediktor menyebabkan taksiran parameter regresi yang dihasilkan memiliki error yang sangat besar. Kasus multikolinieritas dapat dilihat dengan beberapa cara, salah satu cara yang dapat digunakan untuk melihat situasi multikolinieritas adalah melalui nilai VIF. Nilai VIF yang bernilai >10 menunjukkan adanya multikolinieritas pada variabel prediktor[9].

F. Estimasi Parameter Model Regresi Poisson

Estimasi parameter model menggunakan metode MLE (*Maximum Likelihood Estimation*). Parameter model yang ditaksir dengan MLE dinyatakan dengan $\hat{\beta}_p$, dapat diperoleh dengan mencari turunan pertama fungsi \ln *likelihood* dan dilanjutkan dengan mencari turunan kedua terhadap β^T [10] Sebagai berikut.

$$\frac{\partial \ln L(\beta)}{\partial \beta} = - \sum_{i=1}^n x_i \exp(x_i^T \beta) + \sum_{i=1}^n y_i x_i \quad (3)$$

Akan tetapi, penyelesaian dengan metode tersebut sering kali tidak dapat diperoleh hasil yang eksplisit sehingga alternatif lain yang dapat digunakan untuk mendapatkan penyelesaian dari MLE dengan literasi numerik, yaitu Metode *Newton-Raphson*.

G. Pengujian Parameter Model

Salah satu metode yang digunakan dalam penentuan uji statistik dalam pengujian parameter model adalah MLRT (*Maximum Likelihood Ratio Test*). Oleh karena itu, sebelum menentukan uji statistik terlebih dahulu menentukan dua buah fungsi likelihood yang berhubungan dengan model regresi yang diperoleh. Fungsi *likelihood* tersebut adalah $L(\Omega)$ yaitu nilai maksimum likelihood untuk model lengkap dengan melibatkan variabel prediktor. Selain itu, $L(\omega)$ yang merupakan nilai maksimum *likelihood* untuk model sederhana tanpa melibatkan variabel prediktor. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \beta_j \neq 0 ; j = 1, 2, 3, 4$$

Statistik uji yang digunakan adalah sebagai berikut.

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

Keputusan yang diambil akan tolak H_0 apabila $D(\hat{\beta}) > X_{v,\alpha}^2$. Tolak H_0 berarti terdapat salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji yang digunakan adalah sebagai berikut:

$$z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (5)$$

dengan $se(\hat{\beta}_j)$ adalah tingkat kesalahan $\hat{\beta}_j$. Keputusan yang diambil adalah tolak H_0 jika dengan $|z_{hit}| > \frac{z_{\alpha}}{2}$ dengan α adalah tingkat signifikansi.

H. Pengujian Overdispresi

Metode regresi poisson mensyaratkan kondisi dimana nilai *mean* dan varians variabel respon bernilai sama atau kondisi *equidispersion* [11]. Akan tetapi, dalam penerapannya kondisi tersebut sulit dipenuhi oleh data atau sering terjadi *overdispersion* dalam data yang dimodelkan. Kondisi *overdispersion* [12]. Overdispersi akan menyebabkan model dengan estimasi parameter yang bias, selanjutnya dapat menyebabkan kesalahan dalam inferensi untuk parameter tersebut. Oleh karena itu, untuk menguji adanya *overdispersion* pada regresi poisson adalah sebagai berikut.

dengan syarat yang ditunjukkan pada persamaan 6:

$$\theta_1 = \frac{D}{db} > 1; D = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\mu_i} \right\} \quad (6)$$

$$\theta_2 = \frac{\chi^2}{db} > 1; \chi^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i} \quad (7)$$

Overdispersi dapat diketahui melalui nilai devians (D) pada persamaan (5) dan nilai Pearson Chi-Square (χ^2) pada persamaan (6) menghasilkan nilai yang lebih besar dari satu, maka dapat disimpulkan bahwa terjadi overdispersi pada data.

I. Generalized Poisson Regression (GPR)

Generalized Poisson Regression (GPR) dikembangkan untuk menangani asumsi pelanggaran ekuidispersi pada model regresi Poisson. Terutama untuk menyesuaikan model overdispersi serta underdispersi [13].

Model GPR dengan parameter (μ, θ) mirip dengan model regresi Poisson, tetapi diasumsikan bahwa komponen didistribusikan secara acak ke Poisson umum. Dalam analisis GPR, jika θ sama dengan 0 maka modelnya adalah model Poisson. Jika θ lebih dari 0 maka model GPR juga mewakili data yang mengandung overdispersi dan jika θ kurang dari 0 mewakili data yang mengandung underdispersi. Itu menyarankan bahwa Ketika y_i merupakan variabel respon hitungan dan mengikuti distribusi Poisson Umum, probabilitas fungsi kepadatan diberikan bahwa $i = 1, 2, \dots, n$, maka:

$$f(y_i, \lambda_i; \theta) = \left[\frac{\lambda_i}{1+\theta\lambda_i} \right]^{y_i} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} \exp \left[\frac{-\lambda_i(1+\theta y_i)}{1+\theta\lambda_i} \right] \quad (8)$$

Distribusi Poisson tergeneralisasi merupakan perpanjangan alami dari distribusi Poisson. Ketika $\theta = 0$, model dalam persamaan direduksi, bila $\theta > 0$, berarti varians dari distribusi mewakili data hitung dengan overdispersi jika $\theta > 0$, artinya varians lebih kecil dari harapan, yang secara sederhana berarti bahwa distribusi terdapat underdispersi.

J. Estimasi parameter model Generalized Poisson Regression

Estimasi parameter model *Generalized Poisson Regression* dilakukan dengan metode *Maximum Likelihood Estimation*. Fungsi likelihood untuk model GPR adalah sebagai berikut.

$$L(\beta, \theta) = \prod_{i=1}^n \left[\frac{\lambda_i}{1+\theta\lambda_i} \right]^{y_i} \prod_{i=1}^n \frac{(1+\theta y_i)^{y_i-1}}{y_i!} \exp \left[\sum_{i=1}^n \frac{-\lambda_i(1+\theta y_i)}{1+\theta\lambda_i} \right] \quad (9)$$

Estimasi parameter $\hat{\beta}$ diperoleh dengan menurunkan persamaan logaritma natural dari fungsi likelihood terhadap β^T sedangkan estimasi parameter θ diperoleh dengan menurunkan persamaan logaritma natural dari fungsi likelihood terhadap θ dan disamadengankan nol. Penurunan fungsi ln likelihood dilakukan dengan menggunakan metode iterasi Newton-Raphson.

K. Pengujian Parameter Generalized Poisson Regression

Pengujian parameter model *Generalized Poisson Regression* sama dengan pengujian parameter yang digunakan dalam model regresi poisson yaitu dengan metode *Maximum Likelihood Ratio Test* (MLRT). Hipotesis yang digunakan adalah

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \beta_j \neq 0 ; j = 1, 2, 3, 4$$

Statistik uji yang digunakan adalah $D(\hat{\beta})$ pada persamaan (4). Keputusan yang diambil akan tolak H_0 jika $D(\hat{\beta}) > X_{v,\alpha}^2$. Tolak H_0 berarti ada salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji yang digunakan seperti pada persamaan (5) dengan daerah penolakan menolak H_0 jika $|z_{hit}| > \frac{z_{\alpha}}{2}$ dengan α adalah tingkat signifikansi.

L. Negative Binomial Regression

Negative Binomial Regression atau Distribusi Binomial Negatif adalah model non-linear yang berasal dari distribusi *Poisson-gamma mixture* yang merupakan penerapan dari

GLM. Regresi Binomial Negatif digunakan sebagai alternatif dari model regresi poisson yang digunakan untuk mengatasi masalah overdispersi dalam data hitungan [14]. Model regresi Binomial Negatif yang digunakan dalam penelitian ini ditetapkan sebagai berikut:

$$P(y, \mu; \theta) = \frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})y_i!} (\frac{\theta\mu}{1+\theta\mu})^{y_i} (\frac{1}{1+\theta\mu})^{\frac{1}{\theta}} \quad (10)$$

Regresi Binomial Negatif tidak mengasumsikan persamaan *mean* dan varians tetapi mengoreksi overdispersi yang muncul ketika varians lebih besar dari rata-rata.

M. Estimasi Parameter Model Negative Binomial Regression

Metode yang digunakan untuk estimasi parameter dari *Negative Binomial Regression* yaitu metode *Maximum Likelihood Estimate*. Fungsi *likelihood* dari distribusi dinyatakan sebagai:

$$L(\mu, \theta) = \prod_{i=1}^n \frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})y_i!} (\frac{\theta\mu}{1+\theta\mu})^{y_i} (\frac{1}{1+\theta\mu})^{\frac{1}{\theta}} \quad (11)$$

Estimasi parameter $\hat{\beta}$ diperoleh dengan menurunkan persamaan logaritma natural dari fungsi *likelihood* terhadap β^T sedangkan estimasi parameter θ diperoleh dengan menurunkan persamaan logaritma natural dari fungsi *likelihood* terhadap θ dan disamadengankan nol. Penurunan fungsi *ln likelihood* dilakukan dengan menggunakan metode iterasi Newton-Raphson.

N. Pengujian Parameter Model Negative Binomial Regression

Pengujian parameter model *Negative Binomial Regression* sama dengan pengujian parameter yang digunakan dalam model regresi poisson yaitu dengan metode *Maximum Likelihood Ratio Test* (MLRT). Hipotesis yang digunakan adalah dimana θ adalah.

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \beta_j \neq 0 ; j = 1, 2, 3, 4$$

Statistik uji yang digunakan adalah $D(\hat{\beta})$ pada persamaan (4). Keputusan yang diambil akan tolak H_0 jika $D(\hat{\beta}) > X_{v,\alpha}^2$. Tolak H_0 yang berarti ada salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji yang digunakan seperti pada persamaan (5) dengan daerah penolakan menolak H_0 jika $|z_{hit}| > z_{\frac{\alpha}{2}}$ dengan α adalah tingkat signifikansi .

O. Akaike Information Criterion (AIC)

AIC bertujuan untuk mendapatkan model terbaik untuk proses pembangkitan data sebenarnya yang tidak diketahui. Salah satu kriteria informasi yang paling umum digunakan yaitu AIC diperkenalkan oleh Akaike. Idanya adalah untuk memilih model yang meminimalkan kemungkinan jumlah kuadrat galat (Gujarati, 2004). Model regresi terbaik memiliki nilai AIC yang terkecil. Perhitungan nilai AIC dilakukan berdasarkan persamaan berikut [15].

$$AIC = \exp\left(\frac{2k \sum \epsilon_i^2}{n}\right) \quad (12)$$

P. Bayesian Information Criterion (BIC)

BIC berbeda dari AIC hanya dalam istilah pertama yang tergantung pada ukuran sampel n . BIC dirancang untuk menemukan model yang paling mungkin dari data yang diberikan, merupakan salah satu kriteria informasi yang paling banyak digunakan[16]. Tidak seperti AIC, BIC adalah perkiraan faktor Bayes untuk dua model yang digunakan. BIC ditentukan oleh.

$$BIC = -2\ln f(x|\hat{\theta}_k) + p_k \ln n \quad (13)$$

III. METODOLOGI PENELITIAN

Penelitian ini diawali dengan mengambil data COVID-19 di Provinsi Jawa Timur periode Oktober 2020 – Januari 2022. Data yang digunakan merupakan data sekunder. Lalu menentukan variabel Y dan variabel X dari data COVID-19 tersebut. Setelah itu menganalisis statistika deskriptif untuk variabel terikat dan variabel bebas. Kemudian menganalisis korelasi antar variabel-variabel untuk mendeteksi adanya kasus multikolinieritas. Setelah melakukan uji multikolinearitas, dilakukan pemodelan terhadap data untuk mengetahui model terbaik pada regresi poisson. Setelah itu dilakukan uji overdispersi untuk mendeteksi adanya kasus overdispersi pada data dengan melihat nilai *Pearson Chi-squares* dan *Deviance* yang dibagi derajat bebasnya. Jika data terdapat overdispersi, maka data dapat dilanjutkan untuk pemodelan menggunakan *Generalized Poisson Regression* dan *Negative Binomial Regression*.

Selanjutnya adalah melakukan pemodelan data dengan menggunakan *Generalized Poisson Regression*. Setelah mendapatkan model terbaik untuk GPR dengan menaksir parameter GPR, menguji signifikansi parameter secara serentak dan parsial serta menghitung nilai AIC dari model GPR. Lalu dapat dilanjutkan dengan melakukan pemodelan data dengan menggunakan *Negative Binomial Regression*. Setelah mendapatkan model terbaik untuk regresi binomial negatif dengan menaksir parameter, menguji signifikansi parameter secara serentak dan parsial serta menghitung nilai AIC. Pemilihan model regresi terbaik ditentukan berdasarkan nilai AIC dan nilai BIC. Setelah dilakukan pemodelan terhadap data maka dilakukan perbandingan untuk model terbaik hasil GPR dan regresi binomial negatif menggunakan nilai AIC dan nilai BIC.

IV. HASIL DAN PEMBAHASAN

A. Ringkasan Kasus COVID-19

Data yang diambil dari Dinas Kesehatan Jawa Timur sendiri berjumlah 488 data. Dari data diambil empat variabel yang telah disebutkan, data yang diambil sesuai periode penelitian.

Tabel 1 menunjukkan bahwa kasus yang diamati adalah 488 data untuk rata-rata dan standar deviasi kasus kematian harian akibat COVID-19 sebanyak 54.47 kasus dan 84.801 kasus dengan nilai terkecil sebanyak 0 kasus dan nilai terbesar sebanyak 431 kasus. Untuk rata-rata dan standar

deviasi kasus aktif akibat COVID-19 sebanyak 6618.08 kasus dan 11801.784 kasus dengan nilai terkecil sebanyak 91 kasus dan nilai terbesar sebanyak 57126 kasus. Untuk rata-rata dan standar deviasi kasus baru akibat COVID-19 sebanyak 735.93 kasus dan 1264.638 kasus dengan nilai terkecil sebanyak 5 kasus dan nilai terbesar sebanyak 8230 kasus. Untuk rata-rata dan standar deviasi *stringency index* COVID-19 sebanyak 66.5437 persen dan 50.46 persen dengan nilai terkecil sebanyak 50.46 persen dan nilai terbesar sebanyak 72.69 persen. Untuk rata-rata dan standar deviasi *bed occupancy rate* COVID-19 sebanyak 41.99 persen dan 29.1 persen dengan nilai terkecil sebanyak 1.4 persen dan nilai terbesar sebanyak 91.6 persen.

B. Pemodelan Kasus COVID-19 Menggunakan Regresi Poisson

1) Uji Multikolinieritas

Uji Multikolinieritas dilakukan sebagai asumsi awal untuk pendugaan parameter. Kriteria yang dapat digunakan untuk mendeteksi kasus multikolinieritas adalah dengan menghitung nilai VIF.

Nilai VIF masing-masing variabel prediktor disajikan pada Tabel 2 menunjukkan bahwa semua variabel prediktor memiliki nilai VIF < 10. Dengan demikian dapat disimpulkan bahwa semua variabel prediktor tidak terdapat kasus multikolinieritas.

2) Model Regresi Poisson

Pemodelan pada regresi poisson dilakukan dengan cara meregresikan semua kombinasi dari seluruh variabel yang dianalisis. Pengujian parsial dilakukan dengan membandingkan antara nilai *Likelihood Ratio Chi-square* sebesar 34088 dengan nilai $X_{v;\alpha}^2$ sebesar 9,4877 pada model regresi poisson, maka keputusan yang diambil dari uji serentak adalah tolak H_0 , dikarenakan nilai $D(\hat{\beta}) > X_{v;\alpha}^2$. Bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Selanjutnya dilakukan pengujian secara parsial dengan hasil pengujian sebagai berikut.

Mengacu pada Tabel 3 yang menunjukkan bahwa dapat dilihat semua nilai parameter $|Z_{hit}|$ lebih besar dari $|Z_{\alpha/2}|$ sebesar 1.960, sehingga keputusan yang diambil adalah tolak H_0 . Sehingga dapat disimpulkan bahwa parameter berpengaruh signifikan terhadap model. Dari hasil pengujian secara serentak dan juga pengujian parsial tersebut maka model regresi yang dihasilkan sebagai berikut.

$$\hat{\mu} = \exp(-1.657 + 0.000048 X_1 - 0.000029 X_2 + 0.067 X_3 - 0.009 X_4)$$

Dapat dijelaskan bahwa dengan peningkatan setiap 1% pada variabel kasus aktif, *Stringency Index* dan *Bed Occupancy Rate* maka kasus meninggal COVID-19 bertambah masing-masing sebanyak $\exp(0.000048)$ atau sebanyak 1 kasus, $\exp(0.067)$ atau sebanyak 1.07 kasus dan $\exp(0.009)$ atau sebanyak 1 kasus dengan syarat bahwa nilai variabel lain konstan. Sedangkan pada peningkatan setiap 1% pada variabel kasus baru maka jumlah kasus meninggal COVID-19 bertambah sebanyak $\exp(-0.000029)$ atau sebanyak 0.99 kasus dengan syarat bahwa nilai variabel lain konstan. Selain itu dari model regresi tersebut menghasilkan analisis lain menggunakan kriteria AIC dan BIC. Hasil yang

didapatkan kriteria AIC dari estimasi model regresi Poisson sebesar 10217,412 dan didapatkan juga nilai BIC sebesar 10238,363. Dalam pemodelan Regresi Poisson, terdapat asumsi ekuidispersi. Untuk mendeteksi overdispersi, dapat dilihat melalui nilai devians apabila dibagi dengan derajat kebebasan harus lebih besar dari 1.

Tabel 4 menunjukkan bahwa nilai yang didapat deviance/df sebesar 16.298 dan nilai Pearson chi-square/df sebesar 17.378, jika dilakukan uji overdispersi, maka kedua nilai tersebut lebih besar dari 1 sehingga dapat disimpulkan pada model Regresi Poisson tolak H_0 , bahwa pemodelan pada data jumlah kasus meninggal COVID-19 terjadi overdispersi. Maka dapat dilakukan pemodelan dengan menggunakan Regresi Binomial Negatif dan *Generalized Poisson Regression* karena kedua metode tersebut dapat mengakomodasi parameter dispersi.

C. Pemodelan Jumlah Kasus COVID-19 Menggunakan Generalized Poisson Regression

Karena data jumlah kasus meninggal akibat COVID-19 mengalami overdispersi maka model *Generalized Poisson Regression* dapat digunakan. Berikut merupakan hasil untuk pengujian serentak dan pengujian parsial pada model *Generalized Poisson Regression*.

Pengujian parameter secara serentak dilakukan dan berdasarkan nilai *Likelihood Ratio Chi-square* didapatkan sebesar 750.6. Dengan melihat nilai *Likelihood Ratio Chi-square* jika dibandingkan dengan nilai $X_{v;\alpha}^2$ sebesar 9,4877 pada model *Generalized Poisson Regression*, maka keputusan yang diambil adalah tolak H_0 . Dikarenakan semua nilai $D(\hat{\beta}) > X_{v;\alpha}^2$. Sehingga dapat disimpulkan bahwa terdapat seluruh variabel independen yang berpengaruh signifikan terhadap variabel dependen.

Melalui hasil estimasi dengan menggunakan model *Generalized Poisson Regression* seperti yang terlihat pada Tabel 5 yang merupakan hasil estimasi parameter. Selanjutnya dilakukan pengujian parsial untuk model *Generalized Poisson Regression* antara variabel dependen dengan variabel independen. Setelah dilakukannya uji parsial didapatkan hasil bahwa tolak H_0 , dapat dilihat melalui nilai $|Z_{hit}|$ variabel pada Tabel 6 model *Generalized Poisson Regression*. Karena tidak terdapat parameter dengan nilai $|Z_{hit}|$ yang lebih kecil dari $|Z_{\alpha/2}|$. Sehingga dapat ditarik kesimpulan bahwa parameter berpengaruh signifikan terhadap model. Sehingga dapat disimpulkan bahwa semua variabel berpengaruh signifikan terhadap model. Sehingga model yang didapat dari *Generalized Poisson Regression* sebagai berikut.

$$\hat{\mu} = \exp(-1.559 + 1.282 + 0.00005 X_1 - 0.00007 X_2 + 0.037 X_3 + 0.018 X_4)$$

Dapat dijelaskan bahwa dengan peningkatan setiap 1% pada variabel kasus aktif, *Stringency Index* dan *Bed Occupancy Rate* maka kasus meninggal COVID-19 bertambah masing-masing sebanyak $\exp(0.00005)$ atau sebanyak 1 kasus, dan $\exp(0.037)$ atau sebanyak 1.04 kasus, dan $\exp(0.018)$ atau sebanyak 1.02 kasus dengan syarat bahwa nilai variabel lain konstan. Sedangkan pada peningkatan setiap 1% pada variabel kasus baru maka jumlah kasus meninggal COVID-19 bertambah sebanyak $\exp(-$

0.00007) atau sebanyak 0.99 kasus dengan syarat bahwa nilai variabel lain konstan.

Pengujian serentak dan pengujian parsial menunjukkan bahwa semua variabel berpengaruh signifikan terhadap jumlah kasus meninggal akibat COVID-19 di Jawa Timur.

Selain, itu dari model regresi tersebut menghasilkan analisis lain menggunakan kriteria nilai AIC dan nilai BIC. Hasil yang didapatkan kriteria AIC dari estimasi model *Generalized Poisson Regression* sebesar 4061,512 dan didapatkan juga nilai BIC sebesar 4086,654.

D. Pemodelan Kasus COVID-19 Menggunakan *Negative Binomial Regression*

Untuk mengatasi masalah overdispersi dengan pemodelan Regresi Poisson dari data hitungan, juga dapat dilakukan dengan menggunakan model *Negative Binomial Regression*. Sebelum dilakukan pemodelan data menggunakan *Negative Binomial Regression* dilakukan pengujian secara serentak dan pengujian secara parsial.

Dilakukan pengujian parameter secara serentak dengan membandingkan nilai *Likelihood Ratio Chi-square* 697.2 dengan nilai $X^2_{v,\alpha}$ dari alpha (0.05) sebesar 9,4877, maka keputusan yang diambil tolak H_0 . Dikarenakan nilai $D(\hat{\beta}) > X^2_{v,\alpha}$ sehingga dapat disimpulkan bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Selanjutnya dilakukan pengujian parsial terhadap model *Negative Binomial Regression*.

Uji parsial dapat dilihat melalui nilai $|Z_{hit}|$ pada Tabel 6 didapatkan hasil bahwa pada model *Negative Binomial Regression* tolak H_0 karena tidak terdapat parameter dengan nilai $|Z_{hit}|$ yang lebih kecil dari $Z_{\alpha/2}$. Sehingga dapat disimpulkan bahwa semua variabel berpengaruh signifikan terhadap model. Dapat disimpulkan bahwa model yang didapat dari Regresi Binomial Negatif sebagai berikut.

$$\hat{\mu} = \exp(-1.557 + 0.000087 X_1 - 0.00019 X_2 + 0.054 X_3 + 0.02 X_4)$$

Dapat dijelaskan bahwa dengan peningkatan setiap 1% pada variabel kasus aktif, *Stringency Index* dan *Bed Occupancy Rate* maka kasus meninggal COVID-19 bertambah masing-masing sebanyak $\exp(0.000087)$ atau sebanyak 1 kasus, dan $\exp(0.054)$ atau sebanyak 1.05 kasus, dan $\exp(0.02)$ atau sebanyak 1.02 kasus dengan syarat bahwa nilai variabel lain konstan. Sedangkan pada peningkatan setiap 1% pada variabel kasus baru maka jumlah kasus meninggal COVID-19 bertambah sebanyak $\exp(-0.00019)$ atau sebesar 0.99 kasus dengan syarat bahwa nilai variabel lain konstan.

Pengujian serentak dan pengujian parsial menunjukkan bahwa semua variabel berpengaruh signifikan terhadap jumlah kasus meninggal akibat COVID-19 di Jawa Timur.

Selain, itu dari model regresi tersebut menghasilkan analisis lain menggunakan kriteria AIC dan BIC. Hasil yang didapatkan kriteria AIC dari estimasi model *Negative Binomial Regression* sebanyak 4103,979 dan didapatkan juga nilai BIC sebesar 4129,121.

E. Pemilik Model Terbaik

Setelah dilakukan pengujian dari masing-masing model, untuk mengetahui model yang lebih baik dalam pemodelan jumlah kasus meninggal untuk COVID-19 di Jawa Timur

kriteria yang digunakan untuk menentukan model terbaik dipilih berdasarkan perhitungan nilai AIC dan nilai BIC.

Pada Tabel 7 menunjukkan hasil kriteria model untuk ketiga data hitung, yaitu; *Poisson Regression*, *Negative Binomial Regression*, dan *Generalized Poisson Regression*. Hasil penelitian menunjukkan bahwa model yang terpilih adalah *Generalized Poisson Regression* dikarenakan memiliki nilai AIC (4061.512) dan BIC (4086.654) paling kecil diantara ketiga model yang digunakan.

Berdasarkan model jumlah kasus meninggal COVID-19 yang diperoleh dari model *Generalized Poisson Regression* apabila variabel kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* bertambah satu satuan maka jumlah kasus meninggal di Provinsi Jawa Timur akan bertambah dari semula. Jika kasus aktif sebesar 1360 kasus. Kasus baru sebesar 276 kasus *Stringency Index* sebesar 66.2% dan *Bed Occupancy Rate* sebesar 21.95% maka diduga jumlah kasus kematian COVID-19 di Provinsi Jawa Timur bertambah setidaknya sebanyak 14 kasus meninggal.

V. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan pada pemodelan kasus COVID-19 dengan variabel kasus aktif, kasus baru, *Stringency Index*, dan *Bed Occupancy Rate* sebagai faktor yang memengaruhi jumlah korban COVID-19 yang meninggal di Jawa Timur. Hasil yang diperoleh dari analisis menunjukkan bahwa Sebanyak 488 amatan dilakukan dan dihasilkan rata-rata dan standar deviasi pada kasus kematian harian akibat COVID-19 sebanyak 54 kasus dan 85 kasus dengan nilai terkecil sebanyak 0 kasus dan nilai terbesar sebanyak 431 kasus. Dari penelitian ini ditemukan bahwa pengaruh signifikan positif pada variabel kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* terhadap jumlah kematian COVID-19 di Jawa Timur. Hasil analisis menunjukkan pada model Regresi Poisson terdapat adanya overdispersi, sehingga metode *Negative Binomial Regression* dan *Generalized Poisson Regression* dapat digunakan untuk mengatasi overdispersi. Diperlihatkan pula bahwa *Generalized Poisson Regression* adalah model terbaik yang dipilih. Untuk menentukan faktor-faktor yang mempengaruhi kematian akibat COVID-19 di Jawa Timur ketika terdapat indikasi adanya overdispersi pada data hitungan.

Peneliti selanjutnya disarankan kepada Pemerintah Jawa Timur melalui Satuan Tugas COVID-19 pada COVID-19 sebaiknya memperhatikan peningkatan pada kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* karena keempat faktor tersebut merupakan faktor yang dapat meningkatkan kasus kematian terkait COVID-19 di Jawa Timur. Saran untuk penelitian selanjutnya, dapat dilakukan uji autokorelasi pada data jika data merupakan data *timeseries*.

DAFTAR PUSTAKA

- [1] T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Tropical Medicine and International Health*, vol. 25, no. 3, pp. 278–280, Mar. 2020, doi: 10.1111/tmi.13383.
- [2] A. Wilder-Smith, C. J. Chiew, and V. J. Lee, "Can we contain the COVID-19 outbreak with the same measures as for SARS?," *Lancet Infect Dis*, vol. 20, no. 5, pp. e102–e107, 2020, doi: 10.1016/S1473-3099(20)30129-8.

- [3] Y. H. Schukken, G. Casella, and J. van den Broek, "Overdispersion in clinical mastitis in dairy herds: a negative binomial approach," *Prev Vet Med*, vol. 10, no. 3, pp. 239–245, 1991, doi: 10.1016/0167-5877(91)90007-O.
- [4] Kementerian Kesehatan RI, *Pedoman Pencegahan dan Pengendalian Coronavirus Disease (COVID-19)*. Jakarta: Kementerian Kesehatan RI, 2020.
- [5] J. W. Hardin and J. Hilbe, *Generalized Linear Models and Extensions*. Texas: A Stata Press Publication, 2007.
- [6] P. McCullagh, "Generalized linear models," *Eur J Oper Res*, vol. 16, no. 3, pp. 285–292, 1984, doi: 10.1016/0377-2217(84)90282-0.
- [7] P. Nilsson and S. Nilsson, "Application of Poisson Regression on Traffic Safety," Department Mathematical Statistic, Kth Royal Institute Of Technology, Stockholm, 2015.
- [8] R. E. Walpole, *Pengantar Statistika*, 3rd ed. Jakarta: PT Gramedia Pustaka Utama, 1990.
- [9] R. R. Hocking, *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 1st Edition. New York: John Wiley & Sons Inc., 1996.
- [10] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd ed. Hoboken: John Wiley & Sons, 2007.
- [11] T. M. Khoshgoftaar, K. Gao, and R. M. Szabo, "Comparing software fault predictions of pure and zero-inflated Poisson regression models," *Int J Syst Sci*, vol. 36, no. 11, pp. 705–715, Sep. 2005, doi: 10.1080/00207720500159995.
- [12] P. K. T. A. Colin Cameron, *Regression Analysis of Count Data*, 2nd ed. New York: Cambridge University Press, 2013.
- [13] N. Chike and N. Godwin, "Statistical model of road traffic crashes data in anambra state, nigeria: a poisson regression approach," *International Journal Of Scientific & Technology Research*, vol. 4, no. 09, pp. 226–233, 2015.
- [14] F. Famoye, "Restricted generalized poisson regression model," *Commun Stat Theory Methods*, vol. 22, no. 5, pp. 1335–1354, Jan. 1993, doi: 10.1080/03610929308831089.
- [15] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in *Breakthroughs in Statistics: Foundations and Basic Theory*, New York, NY: Springer New York, 1992, pp. 610–624. doi: 10.1007/978-1-4612-0919-5_38.
- [16] S.-P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions," *Accid Anal Prev*, vol. 26, no. 4, pp. 471–482, 1994, doi: 10.1016/0001-4575(94)90038-8.