



Pemodelan Generalized Additive Model For Location, Scale, and Shape (Gamlss) Dengan Pemulusan Locally Estimated Scatterplot Smoothing (Loess) pada Kasus Hiv/Aids Di Jawa Timur

Silvia Tri Wahyuni¹, Tiani Wahyu Utami¹, Moh. Yamin Darsyah²

¹Program Studi Statistika, FMIPA Universitas Muhammadiyah Semarang.

²Program Studi Ilmu Politik, FISIP UIN Walisongo Semarang

* Correspondence: triwahyuni@gmail.com

ABSTRACTS

HIV / AIDS is a contagious disease that can attack all age groups of the population and is a health challenge in almost all over the world including Indonesia. Therefore, it is necessary to model HIV / AIDS cases for the factors that are suspected to influence them. One suitable method for estimating factors that influence HIV / AIDS is the Generalized Additive Model for Location, Scale, and Shape (GAMLSS). The GAMLSS method is flexible because it includes expansion of a good exponential family distribution to handle overdispersion data, continuous data, and discrete data. This research will apply GAMLSS semiparametric modeling with LOESS smoothing to find out the characteristics and models of HIV / AIDS cases in East Java in 2017. Based on the analysis, it was found that the variables that significantly affected were the number of homeless people, number of victims of drug abuse, population poor, and the number of fertile age couples using condom contraception with AIC value of 437,404, degree = 1 and span = 0.3, and the distribution used is Negative Binomial I.

ARTICLE INFO

Article History:

Received 08 Okt 2020

Revised 27 Mei 2021

Accepted 27 Mei 2021

Available online 31 Mei 2021

Keyword:

Generalized Additive Model for Location, Scale, and Shape (GAMLSS), Locally Estimated Scatterplot Smoothing (LOESS), HIV/AIDS

ABSTRAK

HIV/AIDS merupakan salah satu penyakit menular yang dapat menyerang seluruh kelompok umur penduduk dan menjadi tantangan kesehatan hampir diseluruh dunia termasuk Indonesia. Oleh sebab itu, perlu dilakukan pemodelan kasus HIV/AIDS terhadap faktor-faktor yang diduga mempengaruhinya. Salah satu metode yang sesuai untuk menduga faktor-faktor yang mempengaruhi HIV/AIDS adalah Generalized Additive Model for Location, Scale, and Shape (GAMLSS). Metode GAMLSS bersifat fleksibel karena mencakup perluasan dari distribusi keluarga eksponensial yang baik untuk menangani data overdispersi, data kontinu, dan data diskrit. Pada penelitian ini akan menerapkan pemodelan semiparametrik GAMLSS dengan pemulusan LOESS untuk mengetahui karakteristik dan model pada kasus HIV/AIDS di Jawa Timur pada tahun 2017. Berdasarkan hasil analisis, didapatkan bahwa variabel yang berpengaruh secara signifikan adalah jumlah gelandangan, jumlah korban penyalahgunaan Napza, jumlah penduduk miskin, dan jumlah pasangan usia subur menggunakan alat kontrasepsi kondom dengan nilai AIC sebesar 437,404, degree = 1 dan span = 0,3, serta distribusi yang digunakan adalah Negative Binomial I.

Keyword:

Generalized Additive Model for Location, Scale, and Shape (GAMLSS), Locally Estimated Scatterplot Smoothing (LOESS), HIV/AIDS.

1. PENDAHULUAN

HIV atau Human Immunodeficiency Virus adalah sejenis virus yang menyerang atau menginfeksi sel darah putih yang menyebabkan turunnya kekebalan tubuh manusia. AIDS atau Acquired Immune Deficiency Syndrome adalah sekumpulan gejala penyakit yang timbul karena turunnya kekebalan tubuh yang disebabkan infeksi oleh HIV. Akibat menurunnya kekebalan tubuh, maka orang tersebut sangat mudah terserang berbagai penyakit infeksi (infeksi oportunistik) yang sering berakibat fatal (Kemenkes RI, 2014).

Permasalahan HIV dan AIDS menjadi tantangan kesehatan hampir diseluruh dunia termasuk Indonesia. Jumlah kasus HIV di Indonesia cenderung meningkat setiap tahunnya dibandingkan dengan jumlah kasus AIDS. Berdasarkan data WHO (2016), pada tahun 2015 Indonesia menduduki peringkat kedua sebagai penyumbang penderita HIV/AIDS terbanyak di Asia Tenggara. Ditjen P2P Kemenkes RI (2016) menyebutkan bahwa di tahun 2016 Indonesia mengalami kenaikan kejadian insiden HIV menjadi 41.250 orang yang sebelumnya sebesar 30.935 orang pada tahun 2015.

Analisis regresi adalah salah satu metode statistik yang sering digunakan, terutama untuk memodelkan hubungan antara variabel respon dengan variabel prediktor. Penerapan model regresi juga diharapkan semakin fleksibel dengan berbagai jenis data dan algoritma yang memungkinkan untuk menganalisis data secara cepat dan tepat.

Perkembangan regresi diawali oleh Francis Galton (1886) yang memperkenalkan model regresi linier. Kemudian regresi linier berkembang menjadi Generalized Linear Model (GLM) untuk mengatasi variabel respon yang tidak berdistribusi normal namun berdistribusi eksponensial dan memiliki hubungan yang linier. GLM sendiri dapat diperluas menjadi Generalized Additive Model (GAM) agar lebih fleksibel dalam penggunaannya pada kasus tidak ada hubungan linier antara variabel respon dengan variabel prediktor. GAM dan GLM memiliki kekurangan, yaitu tidak dapat memodelkan skewness dan kurtosis secara eksplisit dalam variabel prediktor, melainkan secara implisit melalui ketergantungan skewness (kemencengan) dan kurtosis (keruncingan) terhadap μ .

Berdasarkan kekurangan tersebut, Rigby dan Stasinopoulos, et al., (2007) memperkenalkan suatu metode yang disebut Generalized Additive Model for Location, Scale & Shape (GAMLSS) yang memiliki perluasan GLM dan GAM. GAMLSS dapat memodelkan 4 parameter sekaligus dari

distribusi yang belum bisa dipenuhi oleh GLM dan GAM. Selain itu, metode ini juga lebih fleksibel karena mencakup perluasan dari distribusi keluarga eksponensial untuk menangani data overdispersi, termasuk data dengan kemencengan yang tinggi (highly skew), baik data kontinu maupun diskrit. Sebagai model semiparametrik, GAMLSS merupakan perluasan dari GAM sehingga memiliki fungsi pemulus atau smoothing.

Menurut Rifka (2016), pemulus dibutuhkan dalam memodelkan suatu data dikarenakan kurva yang dihasilkan memiliki nilai estimasi yang besar, sehingga menghasilkan kurva yang kurang halus ataupun terlalu halus. Teknik smoothing yang digunakan adalah Locally Estimated Scatterplot Smoothing (LOESS). LOESS merupakan suatu strategi untuk pemulusan kurva dari data empiris dan menyediakan suatu rangkuman grafis hubungan antara variabel respon dan variabel prediktor (Jacoby, 2000). Prosedur LOESS memungkinkan fleksibilitas yang baik karena tidak ada asumsi tentang bentuk parametrik dari permukaan regresi yang dibutuhkan (SAS Institute Inc, 1999).

Terdapat beberapa penelitian terkait GAMLSS, di antaranya adalah Fauziah (2015) yaitu menganalisis hubungan antara jumlah kematian akibat pneumonia dengan jumlah kasus gizi buruk, cakupan imunisasi dasar yang diberikan, jumlah rumah tangga yang sehat dan bersih, serta jumlah puskesmas setiap provinsi pada tahun 2013 dengan data berupa data diskrit. Hasil penelitian tersebut menyatakan bahwa distribusi yang paling sesuai adalah Binomial Negatif Tipe 2. Penelitian lainnya dilakukan oleh Gilchrist, et al., (2011) yang juga telah menerapkan GAMLSS dengan algoritma RS dan pemulusan LOESS untuk memodelkan parameter distribusi pada data pendapatan film box-office di tahun 1930 yang highly skewness berupa data kontinu. Hasilnya model Box Cox Power Eksponensial (BCPE) adalah model pengepasan terbaik untuk data tersebut.

Berdasarkan penelitian di atas, maka penelitian kali ini akan mengaplikasikan GAMLSS dengan pemulusan LOESS untuk data diskrit. Data yang digunakan dalam penelitian ini adalah data sekunder tahun 2017 dengan jumlah penderita HIV/AIDS sebagai variabel respon (Y), dan jumlah gelandangan, jumlah korban penyalahgunaan Napza, jumlah wanita rawan sosial ekonomi, jumlah penduduk miskin, dan jumlah pasangan usia subur menggunakan alat kontrasepsi kondom sebagai variabel prediktor (X).

Dalam penelitian ini studi kasus yang akan digunakan adalah untuk mengetahui faktor yang

mempengaruhi Indeks Pembangunan Manusia (IPM) menggunakan metode Resampling Bootstrap Residual untuk estimasi parameter model Regresi Linier Berganda.

2. TINJAUAN PUSTAKA

2.1. Model Linier

Bentuk umum regresi linier dapat di tuliskan sebagai berikut (Tirta, 2009):

$$Y = X\beta + \varepsilon \tag{2.1}$$

dengan:

Di mana Y adalah peubah tetap yang bukan acak, β merupakan parameter yang menentukan peubah tetap tadi, dan ε merupakan kesalahan atau galat yang diasumsikan merupakan peubah acak yang berasal dari suatu distribusi tertentu.

Pada model linier normal variabel respon hanya diasumsikan berdistribusi normal, namun dalam kenyataan variabel respon tidak hanya berdistribusi normal saja melainkan juga berdistribusi eksponensial yang dikenal dengan *Generalized Linear Model* (GLM).

Nelder dan Wedderburn mengembangkan model linier yang dikenal dengan *Generalized Linear Model* (GLM) atau model linier tergeneralisir. Model ini mempunyai cakupan distribusi yang lebih luas, yaitu menggunakan asumsi bahwa respon berdistribusi keluarga eksponensial di mana distribusi normal termasuk di dalamnya.

Menurut Nelder dan Wedderburn (1972) GLM terdiri dari 3 komponen, yaitu:

1. Komponen acak, menentukan distribusi bersyarat dari variabel respon, di mana Y_i saling bebas atau independen. Distribusi dari Y_i adalah anggota dari keluarga eksponensial, seperti Distribusi Normal, Binomial, Poisson, Gamma, atau distribusi dari keluarga Invers-Gaussian.
2. Prediktor linier, yaitu sebuah fungsi dari $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$ di mana x_i adalah variabel prediktor untuk unit sebanyak j dengan efek acak β .
3. Fungsi *link*, yaitu fungsi yang mentransformasikan ekspektasi dari variabel respon, $\mu_i = E(Y_i)$ dengan prediktor linier:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \tag{2.2}$$

2.2. Generalized Additive Model (GAM)

GAM adalah suatu generalisasi dari model aditif dan diperkenalkan untuk menyelesaikan masalah yang tidak dapat diselesaikan oleh model aditif. Menurut Hastie dan Tibhsirani (1986), GAM

adalah suatu regresi semi parametrik karena dapat dimodelkan linier, polynomial, dan non parametrik. Model dari GAM dituliskan dalam persamaan 2.3

$$g(\mu_i) = f(X_1, X_2, \dots, X_p) = s_0 + s_1(X_1) + s_2(X_2) + \dots + s_p(X_p) \tag{2.3}$$

Dengan $s_j(X), j = 0, 1, \dots, p$ adalah fungsi *smoothing* (fungsi penghalus).

2.3. Generalized Additive Model for Location and Shape (GAMLSS)

Pada GAMLSS variabel respon berasal dari distribusi keluarga eksponensial dan tambahan distribusi-distribusi lain termasuk untuk distribusi diskrit dan kontinu dengan highly skewed dan kurtosis. Untuk jenis respon cacahan, metode ini cocok untuk data yang mengalami overdispersi dengan menggunakan distribusi overdispersi untuk data diskrit.

2.3.1 Bentuk dan Asumsi GAMLSS

Rigby dan Stasinopoulos (2007) mendefinisikan model dari GAMLSS sebagai berikut. Misalkan $y^T = y_1, y_2, y_3, \dots, y_n$ dengan n adalah panjang vektor dari variabel respon, $k = 1, 2, 3, 4$, dan $g_k(\cdot)$ diketahui sebagai fungsi link monotonik yang menghubungkan antara parameter distribusi dengan variabel prediktor, maka

$$g_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk} \tag{2.4}$$

Jika $Z_{jk} = I_n$, dengan I_n adalah matriks identitas berukuran $n \times n$ dan $\gamma_{jk} = h_{jk} = h_{jk}(x_{jk})$ untuk semua kombinasi dari j dan k , maka didapat bentuk lain dari GAMLSS yang dapat di tuliskan sebagai berikut:

$$g_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} h_{jk} x_{jk} \tag{2.5}$$

$$g_1(\mu) = \eta_1 = X_1 \beta_1 + \sum_{j=1}^{J_1} h_{j1} x_{j1}$$

$$g_2(\sigma) = \eta_2 = X_2 \beta_2 + \sum_{j=1}^{J_2} h_{j2} x_{j2}$$

$$g_3(v) = \eta_3 = X_3 \beta_3 + \sum_{j=1}^{J_3} h_{j3} x_{j3}$$

$$g_4(\tau) = \eta_4 = X_4 \beta_4 + \sum_{j=1}^{J_4} h_{j4} x_{j4}$$

Keterangan:

$\mu, \sigma, v, \tau, \eta_k$ = vektor dengan panjang n

$\beta_k^T = \beta_{1k}, \beta_{2k}, \beta_{3k}, \dots, \beta_{J_k k}$ adalah sebuah vektor parameter

X_k = matriks berukuran $n \times J_k$

h_{jk} = fungsi *smooth* non-parametrik dari variabel prediktor $x_{jk}, J = 1, 2, \dots, j_k$ dan $k = 1, 2, 3, 4$.

di mana x_{jk} untuk $J = 1, 2, \dots, j_k$ juga vektor dengan panjang n . Fungsi h_{jk} adalah fungsi tak diketahui dari variabel prediktor X_k dan h_{jk} adalah sebuah vektor yang mengevaluasi fungsi h_{jk} pada (x_{jk}) .

2.4 Metode Pemulusan Locally Estimated Scatterplot Smoothing (LOESS)

LOESS merupakan akronim dari local regression, yaitu suatu strategi untuk pemulusan kurva dari data empiris dan menyediakan bentuk rangkuman grafis hubungan antara variabel respon dan variabel prediktor (Jacoby, 2000).

Cleveland (1979) menyatakan ada empat pokok yang menjadi dasar untuk melakukan pemulusan LOESS, yaitu pemilihan f (span), W (fungsi bobot), d (derajat polynomial), dan t (iterasi). Parameter f merupakan span (rentang) untuk menentukan jumlah pemulusan yang memberikan takaran dari sebuah observasi yang digunakan setiap daerah regresi. Nilai span di spesifikasi antara 0 sampai 1, yang mana nilai f tersebut mempengaruhi pemulusan terhadap kurva, semakin besar nilai f , maka fitting kurva semakin mulus, dan berlaku untuk sebaliknya.

2.5 Distribusi Data Cacahan pada GAMLSS

Beberapa distribusi data berpotensi untuk data cacahan atau diskrit pada GAMLSS adalah sebagai berikut:

a. Distribusi *Negative Binomial I*

Dalam Distribusi *Negative Binomial Type I* (NBI) diketahui memiliki dua parameter, yaitu μ dan σ . Di mana μ sebagai *mean* dan σ sebagai parameter dispersi bentuk umum dari fungsi kepadatan peluang dari $Y \sim NB(\mu, \alpha)$ diberikan oleh:

$$f(y|\mu, \sigma) = \frac{\Gamma(y+1/\sigma)\alpha^y}{\Gamma(y+1)\Gamma(1/\sigma)} \left[\frac{(\mu\sigma)^y}{(\mu\sigma+1)} \right]^{y+\frac{1}{\sigma}}, \text{ untuk } y = 0, 1, \dots, \infty, \quad (2.6)$$

dengan *mean* $E(y) = \mu$ dan varian $\sigma = \mu(1 + \alpha)$, sehingga saat $\alpha > 0$, varian akan melebihi *mean* dan terjadi overdispersi.

b. Distribusi *Negative Binomial II*

Dalam Distribusi *Negative Binomial Type II* (NBII) atau juga bisa disebut Binomial Negatif Kuadratik, *Negative Binomial Kuadratik* diketahui memiliki dua parameter, yaitu μ dan σ . Menurut Stasinopoulos dan Rigby (2014) fungsi kepadatan peluang dari NBII dapat dituliskan sebagai berikut:

$$f(y|\mu, \sigma) = \frac{\Gamma(y+(\mu/\sigma))\sigma^y}{\Gamma(\mu/\sigma)\Gamma(y+1)(1/\sigma)^{y+(\mu/\sigma)}}, \text{ untuk } y = 0, 1, \dots, \infty, \quad (2.7)$$

dengan $E(y) = \mu$ dan varian $\sigma = \mu + \alpha\mu^2$, sehingga saat $\alpha > 0$, maka akan terjadi overdispersi karena varian lebih besar dari *mean*.

c. Distribusi Poisson Invers Gaussian

Distribusi poisson invers gaussian merupakan salah satu distribusi *mixed poisson*. Bentuk dari distribusi *mixed poisson* tergantung pada distribusi pada random efek (v). Misalkan $g(v)$ adalah fungsi kepadatan peluang dari v dan distribusi marjinal untuk Y diperoleh dengan integral v :

$$P(Y = y|\mu) = \int f(y|\mu, v)g(v)dv \quad (2.8)$$

Untuk distribusi Poisson Invers Gaussian, v diasumsikan mengikuti distribusi invers gaussian dan memiliki fungsi kepadatan peluang yang dapat ditulis sebagai berikut.

$$g(v) = (2\pi\tau v^3)^{-0,5} e^{-(v-1)^2/2\tau v}, v > 0, \quad (2.9)$$

di mana

$$\tau = \text{Var}(V), E(V) = 1$$

Pada akhirnya distribusi PIG dilambangkan dengan $PIG(\mu, \tau)$ diberikan oleh :

$$P(y|\mu, \tau) = \left(\frac{2z}{\pi}\right)^{\frac{1}{2}} \frac{\mu^y e^{\frac{1}{\tau}K_s(z)}}{(z\tau)^y y!}, \quad (2.10)$$

di mana

$$s = y - \frac{1}{2} \text{ dan } z = \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}},$$

$K_s(z) = K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right)$ adalah fungsi

Bessel modifikasi jenis ketiga (Wilmott, 1987).

2.6 Akaike's Information Criterion (AIC)

Akaike's Information Criterion (AIC) adalah metode yang berguna untuk mendapatkan model terbaik yang ditemukan oleh Akaike. AIC memperkirakan kualitas masing-masing model relative terhadap model lain. Misalkan L adalah nilai maksimum dari fungsi *likelihood* suatu model, dan k adalah jumlah parameter yang diestimasi dalam model tersebut, maka nilai AIC dari model tersebut adalah sebagai berikut.

$$AIC = -2\ln(L(\hat{\theta})) + 2k \quad (2.16)$$

Apabila diberikan beberapa model untuk sebuah set data, maka model yang lebih baik adalah model dengan AIC terkecil (Akaike, 1978).

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik Jawa Timur tahun 2018 melalui. Data ini mencakup jumlah penderita kasus HIV/AIDS, jumlah gelandangan, jumlah korban penyalahgunaan Napza, jumlah wanita rawan sosial ekonomi, jumlah penduduk miskin, dan jumlah pasangan usia subur menggunakan alat kontrasepsi kondom berdasarkan 38 kabupaten/kota di Provinsi Jawa Timur. Variabel yang digunakan dalam penelitian ini terbagi menjadi Y dan X dengan Y adalah variabel dependen dan X adalah variabel independen dengan unit penelitian

setiap Kabupaten/Kota di Jawa Timur pada tahun 2017 Adapun langkah-langkah analisis sebagai berikut :

1. Deskripsi data.
2. Menentukan distribusi terbaik berdasarkan model linier.
3. Melihat sebaran data pada *scatterplotmatrix* untuk menentukan variabel yang akan dimodelkan ke dalam parametrik ataupun non parametrik.
4. Pengepasan model semiparametrik GAMLSS dengan pemulusan LOESS.
 - a. Memilih *degree* (derajat) polynomial
 - b. Memilih *span* terbaik
 - c. Menentukan model terbaik
 - d. Uji diagnostik model
5. Menentukan nilai estimasi parameter dari model terbaik.
6. Menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

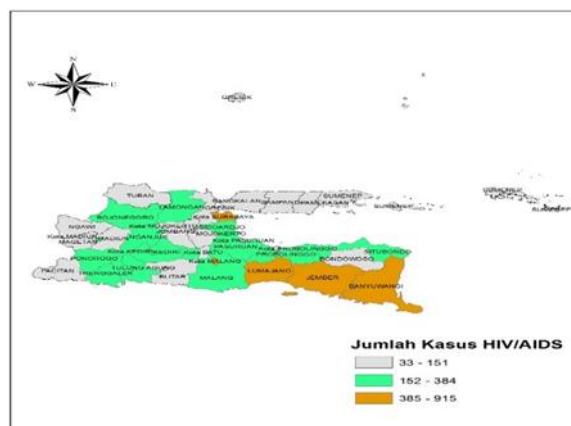
4.1 Deskripsi Data

Tabel 4.1 Deskripsi Jumlah Kasus HIV/AIDS tahun 2017 dan faktor-faktor yang Mempengaruhinya

Var	Min	Mean	Maks	Varian
Y	33	212	915	109,22
X ₁	0	96,16	1.911	3,71
X ₂	0	83,18	541	14,45
X ₃	0	2.899	12.263	88,35
X ₄	7284	117.278	283.961	139,75
X ₅	640	2.961	21.494	345,14

Jumlah kasus HIV/AIDS di Jawa Timur pada tahun 2017 tercatat sebesar 8.056 kasus. Jumlah kasus HIV tertinggi tahun 2017 untuk setiap Kabupaten/kota di Jawa Timur sebesar 915 jiwa. Sedangkan jumlah kasus penyakit HIV/AIDS terendah pada tahun 2017 sebesar 33 jiwa. Unit observasi dari penelitian yaitu 29 Kabupaten dan 9 kota yang ada di Jawa Timur.

Persebaran kasus HIV/AIDS di Jawa Timur berdasarkan kabupaten/kota dapat ditampilkan dalam bentuk peta, hal ini dilakukan dengan tujuan agar persebaran data dapat diketahui dengan mudah. Persebaran data dibagi menjadi 3 kelompok kategori data yakni kategori sangat tinggi, sedang, dan rendah. Warna coklat menunjukkan bahwa persebaran kasus HIV/AIDS sangat tinggi, warna hijau menunjukkan bahwa persebaran kasus HIV/AIDS sedang, dan warna abu-abu menunjukkan bahwa persebaran kasus HIV/AIDS rendah.



Gambar 4.1 Pola Penyebaran Kasus HIV/AIDS di Jawa Timuu Tahun 2017

Jumlah Kasus HIV menurut Kabupaten/kota di Jawa TimurJumlah kasus penyakit HIV/AIDS kategori tinggi dan memiliki angka berkisar antara 385-915 adalah Kota Surabaya, Kota Malang, Lumajang, Jember, dan Banyuwangi. Kabupaten/kota yang memiliki kasus HIV/AIDS kategori sedang dan memiliki angka berkisar antara 152-384 adalah Lamongan, Bojonegoro, Nganjuk, Kota Madiun, Ponorogo, Kediri, Kota Kediri, Kota Mojokerto, Tulungagung, Trenggalek, Sidoarjo, Kota Pasuruan, Pasuruan, Malang, Kota Probolinggo, Probolinggo, dan Situbondo. Sedangkan kabupaten/kota yang memiliki kasus HIV/AIDS kategori rendah dan memiliki angka berkisar antara 33-151 adalah Ngawi, Madiun, Magetan, Tuban, Pacitan, Gresik, Kota Mojokerto, Jombang, Mojokerto, Kota Batu, Blitar, Bangkalan, Sampang, Pamekasan, Sumenep, dan Bondowoso.

4.2 Menentukan Distribusi Terbaik Berdasarkan Model Linier

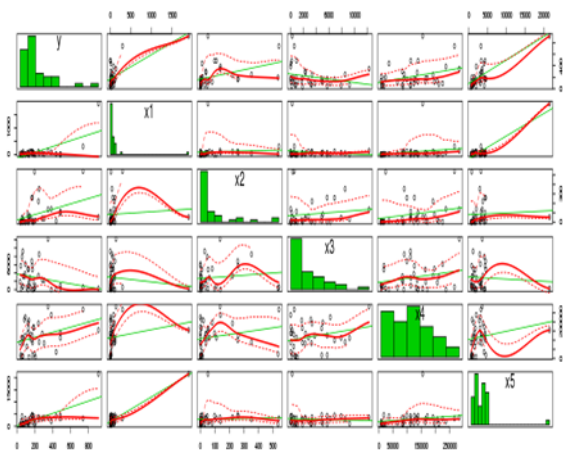
Tabel 4.2 Nilai AIC Model Linier pada Distribusi

Distribusi	AIC	Distribusi	AIC
NB I	447,4158	Log	19317,5
NBII	448,2634	Geom	479,486
Delaporte	12660,41	ZIPIG	448,282
Poisson	1663,867	Yule	780,309
PIG	446,2815	ZAL	41637,5
ZIP	1665,981	ZANB	449,417
ZIP2	1687,415	ZAP	1665,98

Penentuan Distribusi terbaik berdasarkan nilai AIC terkecil. Dalam penelitian ini, diambil 3 distribusi dengan nilai AIC terkecil, yaitu *Negative Binomial I* (NBI), *Negative Binomial II* (NBII), dan *Poisson Invers Gaussian* (PIG).

4.3 Sebaran Data Hubungan Antarvariabel

Untuk melakukan pemodelan GAMLSS dengan pendekatan semiparametrik, perlu dilakukan pemilihan variabel prediktor yang akan dimodelkan dengan parametrik dan semiparametrik dengan melihat sebaran data pada *scatterplotmatrix*.



Gambar 4.2 *Scatterplotmatrix* Hubungan Antara Variabel Respon dan Variabel Prediktor

Berdasarkan hasil *scatterplotmatrix* dihasilkan bahwa variabel prediktor yang akan dimodelkan dalam bentuk parametrik adalah $X_1, X_2, X_4,$ dan X_5 . Sedangkan X_3 dimodelkan dalam bentuk nonparametrik.

4.4 Pengepasan Model Semiparametrik GAMLSS dengan Pemulusan LOESS

4.4.1 Memilih Degree (Derajat) Polinomial

Tabel 4.3 Nilai AIC dengan *Degree* Terbaik

Distribusi	Degree	AIC
Negative Binomial I	1	442,7515
	2	443,0691
Negative Binomial II	1	444,8265
	2	445,7503
Poisson Invers Gaussian	1	443,9205
	2	443,1739

pada distribusi *Negative Binomial I* nilai AIC terkecil untuk pemilihan *degree* terbaik ada pada *degree* = 1, yaitu sebesar 442,7515. Pada distribusi *Negative Binomial II* nilai AIC terkecil untuk pemilihan *degree* terbaik ada pada *degree* = 1, yaitu sebesar 444,8265. Pada distribusi *Poisson Invers Gaussian*, nilai AIC terkecil untuk pemilihan *degree* terbaik ada pada *degree* = 2, yaitu sebesar 443,1739. Setelah pemilihan *degree* terbaik, maka langkah selanjutnya adalah memilih *span* terbaik.

4.4.2 Memilih Span Terbaik

Parameter *span* (rentang) digunakan untuk menentukan berapa banyak jumlah pemulusan yang dibutuhkan untuk memodelkan data, yang mana nilai *span* ini juga mempengaruhi pemulusan terhadap kurva. Menurut (Cleveland, 1979) *span* terbaik ditentukan mulai *span* 0,2 sampai 0,8. Berikut ini merupakan rangkuman nilai AIC dari masing-masing distribusi.

Tabel 4.4 Nilai AIC dengan *Degree* Terbaik

Distribusi	Span	AIC
Negative Binomial I	0,2	438,2214
	0,3	437,4040
	0,4	441,2226
	0,5	444,3035
	0,6	444,4076
	0,7	443,6248
	0,8	442,7787
	Negative Binomial II	0,2
0,3		440,4504
0,4		444,0247
0,5		445,8931
0,6		446,6708
0,7		445,7036
0,8		444,8753
Poisson Invers Gaussian		0,2
	0,3	28239,3045
	0,4	3290,4543
	0,5	437,6684
	0,6	444,4169
	0,7	443,9407
	0,8	443,0789

Pemilihan *span* terbaik yang diperoleh adalah pada distribusi *Negative Binomial I* nilai AIC terkecil untuk pemilihan *span* terbaik ada pada *span* = 0,3, yaitu sebesar 437,4040. Pada distribusi *Negative Binomial II* nilai AIC terkecil untuk pemilihan *span* terbaik ada pada *span* = 0,3, yaitu sebesar 440,4505. Pada distribusi *Poisson Invers Gaussian*, nilai AIC terkecil untuk pemilihan *span* terbaik ada pada *span* = 0,5, yaitu sebesar 437,6684. Setelah pemilihan *span* terbaik

4.4.3 Menentukan Model Terbaik

Pada tahap ini, model terbaik ditentukan dengan melakukan iterasi terhadap model dari setiap distribusi, kemudian membandingkan nilai AIC. Adapun langkah-langkah dalam menentukan model terbaik adalah sebagai berikut.

1. Menginput model yang akan diuji
2. Memasukkan fungsi GAMLSS dengan pemulus LOESS
3. Melakukan pengepasan dengan metode *allpossible elimination* berdasarkan nilai AIC

Tabel 4.5 Nilai AIC dari Model Setiap Distribusi

Distribusi	AIC
Negative Binomial I	437,4040
Negative Binomial II	440,4504
Poisson Invers	437,6683
Gaussian	

Nilai AIC terkecil ada pada distribusi *Negative Binomial I* sebesar 437,404. Oleh karena itu, dapat disimpulkan bahwa distribusi *Negative Binomial I* adalah distribusi yang sesuai dengan model. Adapun ringkasan *coefficient* dan *link function* dari beberapa parameter yang dihasilkan dapat dilihat pada Tabel 4.5 dan Tabel 4.6 berikut ini.

Tabel 4.6 *Coefficient* Parameter μ (mu) dengan Log sebagai *Link function*

	Estimate	Std. Error	t-value	p-value
(Int)	2,045e+00	1,367e-01	14,957	0,000*
X_1	-2,006e-03	4,650e-04	-4,313	0,000*
X_2	2,943e-03	4,238e-04	6,946	0,000*
X_4	1,865e-06	7,711e-07	2,418	0,023*
X_5	3,023e-04	4.289e-05	7,049	0,000*

*) Signifikan dengan taraf 5%

Tabel 4.7 *Coefficient* Parameter σ (sigma) dengan Log sebagai *Link function*

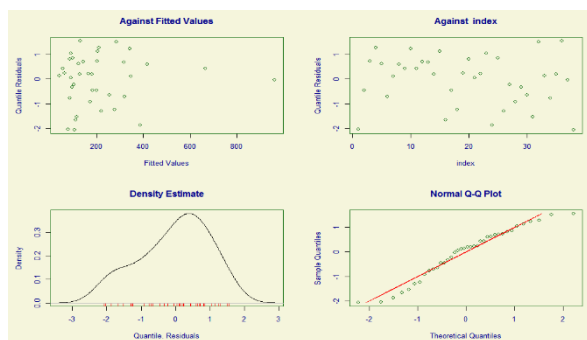
	Estimate	Std. Error	t-value	p-value
(Int)	-2,2437	0,2433	-9,223	0,000*

*) Signifikan dengan taraf 5%

4.4.4 Uji Diagnostik Data

Untuk menentukan kelayakan dari model, maka perlu dilakukan diagnostik data dengan melihat plot residual dari model. Adapun plot residual dari model yang telah didapatkan adalah sebagai berikut:

mean = -0.037614
 variance = 1.016012
 coef. of skewness = -0.43706
 coef. of kurtosis = 2.180344



Gambar 4.3 Plot Residual Model

Suatu model dikatakan mengikuti sebaran normal baku apabila nilai *mean* mendekati 0, nilai *variance* mendekati 1, nilai *skewness* mendekati 0, dan *coefficient of curtosis* mendekati 3 (Stasinopoulus, D.M., and Rigby, R.A. 2015). Jika dilihat pada ringkasan model di atas, maka model dapat didekati dengan sebaran normal. Berdasarkan Gambar 4.4, plot residual dari model terlihat bahwa sebaran Normal Q-Q plot mendekati garis lurus. Hal tersebut juga menunjukkan bahwa model data dapat didekati dengan sebaran normal. Setelah mendapatkan model terbaik, maka selanjutnya adalah menentukan nilai estimasi parameter dari model terbaik.

4.4.5 Menentukan Nilai Estimasi Parameter dari Model Terbaik

Tabel 4.8 Nilai Estimasi dan *p-value*

	Estimate	p-value
(Int)	2,045e+00	0,000*
X_1	-2,006e-03	0,000*
X_2	2,943e-03	0,000*
X_4	1,865e-06	0,023*
X_5	3,023e-04	0,000*

*) Signifikan dengan taraf 5%

Tabel 4.8 memperlihatkan bahwa *p-value* untuk masing-masing variabel prediktor bernilai kurang dari 5%. Sehingga didapatkan hasil bahwa variabel $X_1, X_2, X_4,$ dan X_5 pada model *Negative Binomial I* berpengaruh signifikan terhadap kasus HIV/AIDS di Jawa Timur. Sehingga model yang di hasilkan sebagai berikut:

$$\hat{\mu}_i = \exp (2,045 - 2,006 \times 10^{-3} X_1 + 2,943 \times 10^{-3} X_2 + 1,865 \times 10^{-6} X_4 + 3,023 \times 10^{-4} X_5)$$

Model tersebut menunjukkan bahwa setiap penambahan satu gelandangan (X_1) akan menyebabkan nilai harapan jumlah penderita HIV/AIDS menurun sebesar $\exp(-2,006 \times 10^{-3}) = 0,998$ kali. Artinya setiap penambahan 10000 penduduk miskin akan menurunkan jumlah penderita HIV/AIDS sebanyak 9980 orang dengan asumsi peubah lain dianggap tetap. Kemudian, setiap penambahan Korban Penyalahgunaan Napza (X_2) akan menyebabkan nilai harapan jumlah penderita HIV/AIDS meningkat sebesar $\exp(2,943 \times 10^{-3}) = 1,003$ kali. Artinya setiap penambahan 10000 Korban Penyalahgunaan Napza akan menaikkan jumlah penderita HIV sebanyak 10030 orang. Kemudian, setiap penambahan Penduduk Miskin (X_4) akan menyebabkan nilai harapan jumlah penderita HIV/AIDS meningkat sebesar $\exp(1,865 \times 10^{-6}) = 1$ kali. Artinya setiap

penambahan 10000 Penduduk Miskin akan menaikkan jumlah penderita HIV/AIDS sebanyak 10000 orang. Dan terakhir setiap penambahan Pasangan Usia Subur menggunakan Alat Kontrasepsi Kondom (X_5) akan menyebabkan nilai harapan jumlah penderita HIV/AIDS meningkat sebesar $\exp(3,023 \times 10^{-4}) = 1$ kali. Artinya setiap penambahan 10000 Pasangan Usia Subur menggunakan Alat Kontrasepsi Kondom akan menaikkan jumlah penderita HIV/AIDS sebanyak 10000 orang.

Untuk parameter *scale* (σ), diperoleh hasil sebagai berikut.

$$\hat{\sigma} = \exp(-2,2437)$$

Di mana estimasi nilai dari parameter *scale* adalah $\exp(-2,2437) = 0,106$.

5. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, dapat diambil kesimpulan sebagai berikut.

1. Karakteristik dari jumlah kasus HIV/AIDS di Jawa Timur tahun 2017 dapat dilihat pada tabel deskripsi data dan peta persebaran data, di mana jumlah kasus HIV/AIDS di Jawa Timur pada tahun 2017 tercatat sebesar 8.056 kasus. Jumlah kasus HIV/AIDS kategori tinggi pada tahun 2017 terjadi di 5 kabupaten/kota di Jawa Timur. Jumlah kasus HIV/AIDS kategori sedang pada tahun 2017

6. ACKNOWLEDGMENT

Ucapan Terimakasih kami ucapkan kepada Dinas Kesehatan Provinsi Jawa Timur atas data data yang bisa kami akses serta Laboratorium Statistik Kesehatan UNIMUS.

7. REFERENCES

- Akaike, H. 1978. *A Bayesian Analysis of The Minimum AIC Procedure*. Annals of The Institute of Statistical Mathematics, Part A Hal.914. <http://www.ism.ac.jp/editsec/aism/pdf/>
- Cleveland, W.S. 1979. *Robust Locally Weighted Regression and Smoothing Scatterplots*. Journal of The American Statistical Association, **74**:829-836.
- Cleveland W.S, Devlin, S.J. 1988. An Approach to Regression Analysis by Local Fitting. *Journal of The American Statistical Association*, **83**:596-610.
- Badan Pusat Statistik. 2018. *Jawa Timur Dalam Angka. 2018*. BPS Jawa Timur. Surabaya.
- Departemen Kesehatan RI: *Penanggulangan HIV/AIDS di Indonesia, Respon Saat ini Menangkal Ancaman Bencana Nasional AIDS Mendatang*. Departemen Kesehatan RI. Jakarta: 2002.
- Direktorat Jenderal Pencegahan dan Pengendalian Penyakit Kementerian Kesehatan RI: *Laporan Perkembangan HIV/AIDS dan Penyakit Infeksi Menular Seksual Tahun 2016*.
- Fauziah, L. 2015. *Aplikasi GAMLSS dengan Pemulusan Loess dan Algoritma Rigby-Stasinopoulos Pada Data Cacahan*. Skripsi. Jember: Universitas Jember.
- Hasanah, U. 2015. *Aplikasi GAMLSS dalam Pemodelan Generalized Gamma dengan Algoritma Mixed pada Pemulusan Loess*. Skripsi. Jember: Universitas Jember.
- Hastie, T.J, Tibshirani, R.J. 1990. *Generalized Additive Models*. Chapman and Hall: London.
- Hardin, J. W., dan Hilbe, J. M. 2007. *Generalized Linear Models and Extensions Second Edition*. Texas: Stata Press.
- Hilbe, J.M. 2011. *Negative Binomial Regression Second Edition*. New York: Cambridge University Press.

terjadi di 17 kabupaten/kota di Jawa Timur. Sedangkan jumlah kasus HIV/AIDS kategori rendah terjadi di 16 kabupaten/kota di Jawa Timur. Unit observasi dari penelitian yaitu 29 kabupaten dan 9 kota yang ada di Jawa Timur.

2. Berdasarkan hasil analisis diperoleh model terbaik GAMLSS menggunakan distribusi *Negative Binomial I* untuk bagian parametrik dan LOESS dengan *degree* = 1 dan *span* = 0,3 untuk bagian nonparametrik, serta nilai AIC sebesar 437,4040. Persamaan model tersebut adalah sebagai berikut.

Untuk parameter *location* (μ) diperoleh hasil sebagai berikut.

$$\begin{aligned} \hat{\mu}_i = \exp & (2,045 - 2,006 \times 10^{-3} X_1 \\ & + 2,943 \times 10^{-3} X_2 \\ & + 1,865 \times 10^{-6} X_4 \\ & + 3,023 \times 10^{-4} X_5) \end{aligned}$$

Untuk parameter *scale* (σ), diperoleh hasil sebagai berikut.

$$\hat{\sigma} = \exp(-2,2437)$$

Pemodelan kasus HIV/AIDS di Jawa Timur dengan menggunakan GAMLSS dengan pemulusan LOESS diperoleh bahwa faktor-faktor yang berpengaruh adalah jumlah gelandangan, korban penyalahgunaan Napza, jumlah penduduk miskin, dan jumlah pasangan usia subur menggunakan alat kontrasepsi kondom.

- Herindrawati., dkk. 2017. *Pemodelan Regresi Poisson Inverse Gaussian Studi Kasus : Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015*. Jurusan Statistika. Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS).
- Jacoby, W.G. 2000. *Loess : a nonparametric, graphical tool for depicting relationships between variables*. Electoral Studies.**19**:577-613.
- Kemendes RI. 2006. *Profil Kesehatan Indonesia Tahun 2005*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Kemendes RI. 2014. *Profil Kesehatan Indonesia Tahun 2013*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Nelder, J.A.dan Wedderburn, R. W. M. 1972. *Generalized Linear Models*, *J. R Statist. Soc. A.*,**135**: 370-384
- Pasokawati, Tsamara. 2019. *Pemodelan Geographically Weighted Negative Binomial Regression Pada Kasus HIV di Jawa Timur*. Skripsi. Semarang: Universitas Muhammadiyah Semarang.
- Ramachandran, K. M dan Tsokos, C.P. 2009. *Mathematical Statistic with Application*. United States of America: Academic Press.
- Rigby,R.A dan Stasinopoulos D.M.2005 Generalized Additive Models for Location, and Shape. *Applied Statistic*,**54**.507-554
- Rohimah, S., R. 2015. *Model Spasial Autoregresif Poisson Untuk Mendeteksi Faktor-Faktor Yang Berpengaruh Terhadap Jumlah Penderita HIV Di Provinsi Jawa Timur*. Program Studi Matematika, Jurusan Matematika, FMIPA, Universitas Negeri Semarang.
- Stasinopoulos, D. M. & Rigby, R.A. 2007. *Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R*. J.R.Statisc.
- Stasinopoulos,D.M., Rigby,R.B.,Akantziliotou,C.2008. *Instruction on How to Use The GAMLSS Package in R Second Edition*. STORM Research Centre, London Metropolitan University, London. <http://www.gamlss.com/>
- SAS Institute Inc. 1999. SAS Campus Drive, Cary, North Carolina 27513.
- Tirta, IM. 2009. *Analisis Regresi dengan R (ANRER)*. Jember: UPT Penerbitan Universitas Jember
- Wahyuni, Rifka. 2016. *Model Aditif Tergeneralisir Semiparametrik dengan Menggunakan Pemulusan Loess pada Kasus Pendugaan Curah Hujan*. Skripsi. Jember: Universitas Jember.
- World Health Organization. (2016). *HIV and Young People Who Inject Drugs*. Geneva: WHO Document Production Services.