Seminar Nasional Sains Data 2025 (SENADA 2025)
UPN "Veteran" Jawa Timur
E-ISSN 2808-5841
P-ISSN 2808-7283

Regresi Nonparametrik Kernel *Triangle* dan *Gaussian* dengan Estimator *Priestley-Chao* untuk Pemodelan Kasus AIDS di Pulau Jawa

Alfidha Rahmah[1], Tiani Wahyu Utami[2], Alwan Fadlurohman[3]
[1, 2]*Statistika. Universitas Muhammadiyah Semarang*
[1]alfidharhm@gmail.com
[2]tianiunimus@gmail.com
[3]*Sains Data, Universitas Muhammadiyah Semarang*
[3]alwanr21@gmail.com
Corresponding author email: alfidhrhm@gmail.com

Abstract: *Infection with the Human Immunodeficiency Virus (HIV). In Indonesia, the number of AIDS cases increased in 2023, with the highest cases recorded in provinces on the island of Java. This study aims to model the relationship between socioeconomic factors and the number of AIDS cases in Java using the Priestley-Chao kernel nonparametric regression. The nonparametric approach was chosen for its flexibility and because it does not require assumptions about the form of the relationship. The Priestley-Chao estimator has the advantage of producing smooth estimates even with limited data. This study compares two kernel functions: the triangle kernel, which performs well in modeling fluctuating data, and the Gaussian kernel, which is defined over all real numbers. The analysis showed that the best triangle kernel model was obtained with a bandwidth combination of h1 = 0.10, h2 = 0.20, and h3 = 0.25, resulting in a minimum GCV value of 0.003178. The best Gaussian kernel model was found at h1 = 0.10, h2 = 0.14, and h3 = 0.10, with a minimum GCV value of 0.008402. Based on MSE and R^2 values, the triangle kernel demonstrated better performance (MSE = 0.003129; R^2 = 0.887730) compared to the Gaussian kernel (MSE = 0.008247; R^2 = 0.704080).*
Keywords: *Nonparametric Regression, Kernel Triangle, Kernel Gaussian, Priestley-Chao, AIDS*

Abstrak: *Acquired Immunodeficiency Syndrome* (AIDS) merupakan masalah kesehatan global yang disebabkan oleh infeksi *Human Immunodeficiency Virus* (HIV). Di Indonesia, jumlah kasus AIDS meningkat pada tahun 2023 dengan kasus tertinggi tercatat di provinsi-provinsi di Pulau Jawa. Penelitian ini memodelkan hubungan antara faktor-faktor sosial ekonomi dengan jumlah kasus AIDS di Pulau Jawa menggunakan regresi nonparametrik kernel *Priestley-Chao*. Pendekatan nonparametrik dipilih karena sifatnya yang fleksibel dan tidak memerlukan asumsi bentuk hubungan tertentu. Estimator *Priestley-Chao* memiliki keunggulan dalam menghasilkan estimasi yang halus meskipun data terbatas. Penelitian ini membandingkan dua fungsi kernel, yaitu kernel *triangle* yang unggul dalam memodelkan data fluktuatif, serta fungsi kernel *gaussian* yang terdefinisi pada seluruh bilangan riil. Hasil analisis menunjukkan bahwa model kernel *triangle* terbaik diperoleh pada kombinasi *bandwidth* h_1=0.10, h_2=0.20, dan h_3=0.25 dengan nilai GCV minimum 0.003178. Model kernel *gaussian* terbaik diperoleh pada h_1=0.10, h_2=0,14, dan h_3=0,10 dengan GCV minimum 0.008402. Berdasarkan nilai MSE dan R^2, kernel *triangle* memiliki kinerja lebih baik (MSE = 0.003129; R^2=0.887730) dibandingkan kernel *gaussian* (MSE = 0.008247; R^2=0.704080).
Kata kunci: Regresi *Nonparametrik*, Kernel *Triangle*, Kernel *Gaussian*, *Priestley-Chao*, AIDS

I. PENDAHULUAN

Acquired Immunodeficiency Syndrome atau AIDS merupakan masalah kesehatan yang masih menjadi perhatian utama di seluruh dunia. AIDS merupakan sekumpulan gejala atau penyakit di akibatkan rusaknya sistem kekebalan tubuh yang disebabkan oleh penularan HIV [1]. Infeksi virus HIV dapat menyebabkan turunnya kemampuan imunitas tubuh dalam melawan benda-benda asing di dalam tubuh, sehingga dapat menyebabkan AIDS pada tahap terminal infeksinya [2]. Hingga kini, penularan HIV terus berlanjut di seluruh negara dan telah merenggut sekitar 40,4 juta nyawa sejak awal munculnya kasus AIDS [3].

Salah satu strategi utama dalam mengendalikan dan mencegah HIV yaitu dengan mencapai target *triple* 95s atau 95-95-95 di tahun 2030. Target tersebut terdiri dari 95% ODHIV mengetahui statusnya, 95% ODHIV mendapatkan pengobatan *Antiretroviral Therapy* dan 95% ODHIV dalam *Antiretroviral Therapy* mengalami penekanan virus [4]. Namun, capaian Indonesia masih sangat jauh dalam mewujudkan target tersebut. Berdasarkan data laporan perkembangan HIV/AIDS Indonesia tahun 2023

yang dapat dilihat pada gambar 1.1, tercatat jumlah ODHIV hidup dan mengetahui statusnya sebanyak 77% dan hanya 45% yang mendapatkan pengobatan, sedangkan jumlah ODHIV yang sedang mendapatkan pengobatan dan virusnya tersupresi adalah sebanyak 46% [5].

Pada tahun 2023, kasus baru AIDS di Indonesia meningkat signifikan menjadi 16,410 kasus, dibandingkan 9.901 kasus pada tahun 2022 [6]. Provinsi dengan jumlah kasus tertinggi adalah Jawa Barat (2,575 kasus atau 16% dari total nasional), diikuti oleh Jawa Timur (2,432 kasus) dan Jawa Tengah (2,008 kasus), serta beberapa provinsi lain seperti Bali, DKI Jakarta, Sumatera Utara, Banten, Nusa Tenggara Timur, dan Kalimantan Barat [7]. Data ini menunjukkan bahwa kasus AIDS terbanyak masih didominasi oleh provinsi-provinsi di Pulau Jawa.

Upaya pengendalian HIV/AIDS menghadapi tantangan tidak hanya dari akses pengobatan, tetapi juga faktor sosial-ekonomi seperti pengangguran, kemiskinan, dan ketahanan pangan yang meningkatkan risiko perilaku berbahaya dan kerentanan terhadap HIV ([8]. Beberapa penelitian terdahulu juga mengungkapkan bahwa aspek sosial ekonomi mempunyai pengaruh terhadap banyaknya kasus HIV/AIDS di Indonesia diantaranya adalah Indeks Pembangunan Manusia (IPM), persentase penduduk miskin, dan kepadatan penduduk [9-11].

Berdasarkan uraian di atas, analisis terhadap faktor-faktor sosial-ekonomi kasus HIV/AIDS menjadi sangat penting untuk mendapatkan wawasan lebih dalam terkait pola hubungan antar variabel. Salah satu pendekatan yang dapat diterapkan dalam analisis tersebut adalah dengan metode analisis regresi. Analisis regresi merupakan metode statistik yang digunakan untuk mengestimasi hubungan antara variabel-variabel, yaitu dengan memodelkan dan menganalisis bentuk hubungan antara satu variabel dependen dengan satu atau lebih variabel independen [12].

Pada analisis regresi terdapat tiga pendekatan utama, diantaranya parametrik, semiparametrik, dan nonparametrik. Apabila asumsi tidak terpenuhi pada metode parametrik, maka metode nonparametrik dapat digunakan untuk menduga kurva regresi [13]. Pada metode nonparametrik, diharapkan data dapat menemukan estimasi kurva regresi sendiri, tanpa adanya pengaruh oleh elemen subjektif dari perancang penelitian [14]. Berbagai pendekatan dari regresi nonparametrik telah dikembangkan untuk mengestimasi fungsi regresi, salah satunya adalah metode kernel. Menurut [15] estimator kernel memiliki bentuk yang fleksibel serta mudah dikerjakan secara matematik, selain itu juga memiliki rata-rata kekonvergenan yang cenderung cepat.

Penghalusan (*smoothing*) dengan pendekatan regresi nonparametrik kernel sangat bergantung pada pemilihan *bandwidth* dan fungsi kernel [16]. Dalam regresi nonparametrik kernel terdapat beberapa pilihan fungsi kernel, diantaranya adalah *uniform, triangle, epanechnikov, gaussian*, kuartik, dan cosinus [17]. Fungsi-fungsi tersebut memiliki bentuk yang berbeda-beda, di mana masing-masing memiliki karakteristik yang mempengaruhi cara model menyesuaikan diri dengan data. Salah satu fungsi yang umum digunakan adalah fungsi *triangle,* karena fungsi ini mudah dalam perhitungan dan teliti dalam memodelkan data yang fluktuatif. Selain itu, fungsi *gaussian* juga sering digunakan, karena perhitungannya yang sederhana dan terdefinisi pada seluruh bilangan riil [18].

Terdapat beberapa penelitian terdahulu yang menerapkan metode regresi nonparametrik kernel. Penelitian oleh [19] menerapkan regresi nonparametrik kernel dalam mengestimasi data rata-rata bulanan *sunspot*. Penelitian tersebut memperoleh hasil bahwa estimator kernel *triangle* memberikan hasil yang lebih baik daripada kernel *epanechnikov*. Penerapan metode regresi nonparametrik kernel juga dilakukan oleh Handayani (2019) pada jumlah hotel bintang di Indonesia dan faktor pengaruhnya pada tahun 2016. Berdasarkan nilai MSE terkecil yaitu sebesar 4.6581×10^{-4} diperoleh bahwa kernel *gaussian* menghasilkan model yang lebih baik dari kernel kuartik [20]. Penelitian pada [21] menerapkan metode regresi nonparametrik kernel *Priestley-Chao* dengan fungsi *gaussian* dalam mengestimasi nilai neraca perdagangan Indonesia. Hasil pemodelan memperoleh nilai MSE sebesar 0.98 dan R^2 sebesar 86%. Penelitian lain dilakukan oleh [22] dalam memodelkan inflasi berdasarkan tiga faktor pengaruhnya menggunakan regresi nonparametrik kernel dengan fungsi *triangle* dan penaksir *Priestley-Chao*. Hasil penelitian tersebut memperoleh koefisien determinasi sebesar 80%.

Berdasarkan uraian yang telah dijelaskan, penelitian ini akan melakukan pemodelan kasus AIDS berdasarkan faktor-faktor sosial-ekonomi diantaranya persentase penduduk miskin, IPM, dan kepadatan penduduk menggunakan regresi nonparametrik kernel dengan estimator *Priestley-Chao*. Penelitian ini berfokus pada wilayah di Pulau Jawa sebagai wilayah dengan jumlah kasus AIDS yang tinggi. Fungsi kernel yang digunakan adalah *triangle* dan *gaussian*. Penelitian-penelitian sebelumnya menunjukkan bahwa fungsi kernel *triangle* dan *gaussian* terbukti menghasilkan model yang baik, sehingga dalam penelitian ini akan dilakukan perbandingan dari kedua fungsi tersebut. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam memahami dinamika penyebaran AIDS di wilayah Pulau Jawa melalui pendekatan pemodelan yang lebih mendalam.

II. METODE PENELITIAN

2.1. Sumber dan Struktur Data

Dalam penelitian ini menggunakan data kuantitatif yang merupakan data sekunder dari publikasi Provinsi dalam Angka 2024 pada Provinsi Jawa Tengah, Jawa Barat, Jawa Timur, Banten, DI Yogyakarta, dan DKI Jakarta yang diakses pada website resmi Badan Pusat Statistik, yaitu https://www.bps.go.id/id.

Variabel yang digunakan dalam penelitian ini adalah data per kabupaten/kota di Pulau Jawa tahun 2023 yang terdiri dari kasus baru AIDS (Y), Persentase Penduduk Miskin (X1), Indeks Pembangunan Manusia (X2), dan Kepadatan Penduduk (X3) sebagaimana yang disajikan dalam tabel 1.

Tabel 1. Variabel Penelitian

Simbol	Variabel	Skala	Satuan
Y	Jumlah Kasus Baru AIDS per Kabupaten/Kota di Pulau Jawa tahun 2023	Rasio	Kasus
X_1	Persentase Penduduk Miskin per Kabupaten/Kota di Pulau Jawa tahun 2023	Rasio	Persen (%)
X_2	Indeks Pembangunan Manusia per Kabupaten/Kota di Pulau Jawa tahun 2023	Interval	-
X_3	Kepadatan Penduduk per Kabupaten/Kota di Pulau Jawa tahun 2023	Rasio	Jiwa/km^2

2.2. Langkah Penelitian

langkah- langkah yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan data
2. Melakukan analisis statistik deskriptif pada variabel penelitian
3. Melakukan proses normalisasi data menggunakan persamaan *min-max normalization* seperti dalam (1) [23]:

$$z_i^* = \frac{z_i - \min(z)}{\max(z) - \min(z)} \qquad (1)$$

dimana:

z : data sebenarnya
z_i : data ke-i
z_i^* : hasil normalisasi ke-i
min(z) : nilai terkecil data
max(z) : nilai tertinggi data

4. Menentukan *bandwidth* yang akan diuji dalam estimator kernel
5. Pemodelan menggunakan fungsi kernel *triangle* dengan estimator *Priestley-Chao*:
   a. Memasukkan *bandwidth* ke dalam estimator *triangle*
   b. Menghitung nilai *bandwidth* optimum menggunakan GCV seperti dalam (2) [24]:

E-ISSN 2808-5841
P-ISSN 2808-7283

$$\text{GCV} = \frac{\text{MSE}}{\left(\frac{1}{n}tr(\boldsymbol{I}-\boldsymbol{H}(h))\right)^2} \quad (2)$$

dimana :

$MSE = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$

H(h) : $\boldsymbol{X}(\boldsymbol{X}\boldsymbol{X}'\boldsymbol{X}+nh\boldsymbol{I})^{-1}\boldsymbol{X}'$

tr : trace

n : banyaknya data

$\boldsymbol{X}$: matriks data

$\boldsymbol{X}'$: transpos matriks

I : matriks identitas

h : bandwidth

y_i : nilai variabel respon pada observasi ke-i, dengan i =1,2,..., n

$\hat{y}_i$: nilai prediksi dari variabel y_i

c. Menentukan *bandwidth* optimum berdasarkan nilai GCV minimum

d. Menghitung estimasi kasus AIDS dengan estimator *triangle* seperti dalam (3) menggunakan *bandwidth* optimum yang diperoleh [22]:

$$\hat{m}(x) = \sum_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{x_{ji}-x_{j(i-1)}}{h_j}\left(\left(1-\left|\frac{x_j-x_{ji}}{h_j}\right|\right)\text{I}\right)\right)y_i \quad (3)$$

dimana:

$I(x) = \begin{cases}1 \text{ jika } |u| \le 1\\ 0 \text{ jika } |u| > 1\end{cases}$

$\hat{m}(x)$: fungsi estimasi regresi nonparametrik yang melibatkan variabel prediktor

n : jumlah data

x : variabel acak

x_i : nilai variabel prediktor pada observasi ke-i, dengan i =1,2,..., n

h : parameter penghalus atau *bandwith*

e. Melakukan proses denormalisasi hasil estimasi ke skala data asli dengan transformasi persamaan (1) menjadi:

$$z_i = z_i^*(max(z) - \min(z)) + \min(z)$$

6. Pemodelan menggunakan fungsi kernel *gaussian* dengan estimator *Priestley-Chao*:
   a. Memasukkan *bandwidth* ke dalam estimator *gaussian*
   b. Menghitung nilai *bandwidth* optimum menggunakan GCV seperti dalam (2)
   c. Menentukan *bandwidth* optimum berdasarkan nilai GCV minimum
   d. Menghitung estimasi kasus AIDS dengan estimator *gaussian* seperti dalam (4) menggunakan *bandwidth* optimum yang diperoleh:

$$\hat{m}(x) = \sum_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{x_{ji}-x_{j(i-1)}}{h_j}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_j-x_{ji}}{h_j}\right)^2\right)\right)\right)y_i \quad (4)$$

   e. Melakukan proses denormalisasi hasil estimasi ke skala data asli

7. Membandingkan kebaikan model kernel *triangle* dan *gaussian* berdasarkan nilai MSE dan R^2 dengan persamaan (5) dan persamaan (6) [25]:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad (5)$$

dimana:

n : banyaknya observasi

y_i : nilai variabel respon pada observasi ke-i, dengan i =1,2,..., n

$\hat{y}_i$: nilai prediksi dari variabel y_i

Semakin kecil nilai MSE, menunjukkan semakin baiknya suatu model.

E-ISSN 2808-5841
P-ISSN 2808-7283

$$R^2 = 1 - \frac{JKG}{JKT} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (6)$$

dimana:
JKG : jumlah kuadrat galat yang mengukur kesalahan prediksi model
JKT : jumlah total kuadrat yang mengukur total variasi dalam data
y_i : nilai peubah respon ke-i
$\hat{y}_i$: nilai prediksi dari variabel
$\bar{y}_i$: rata-rata peubah respon
Tabel 2 menunjukkan kategori dari nilai R^2 [26].

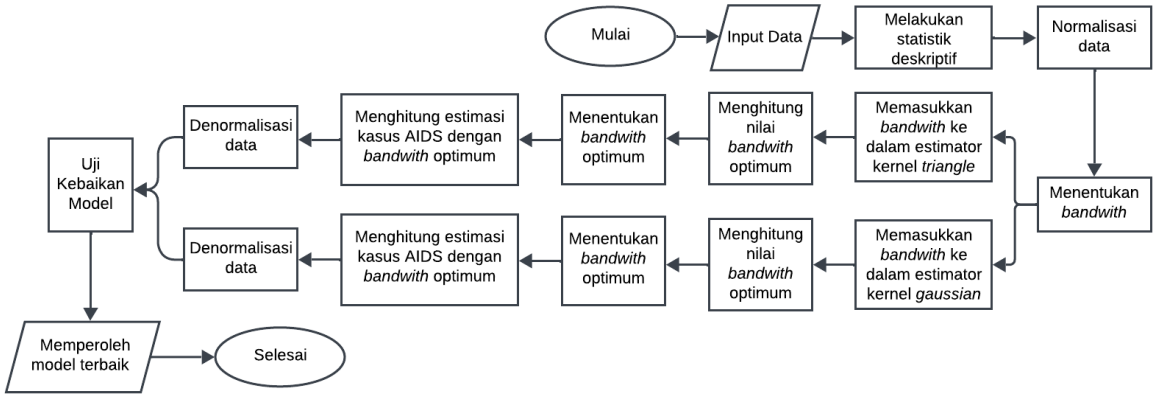
Tabel 2. Kategori Nilai $\mathbf{R^2}$

Nilai	Kategori
$R^2 > 0{,}9$	Sangat Akurat
$0{,}8 \leq R^2 < 0{,}9$	Akurat
$0{,}5 \leq R^2 < 0{,}8$	Cukup Akurat
$R^2 < 0{,}5$	Tidak Akurat

8. Memperoleh fungsi kernel terbaik berdasarkan nilai MSE terkecil dan R^2 terbesar
9. Memperoleh model terbaik dalam mengestimasi kasus AIDS di Pulau Jawa
10. Kesimpulan

2.3. Diagram Alir Penelitian

Diagram alir penelitian pada penelitian ini disajikan pada gambar 1.



Gambar 1. Diagram Alir Penelitian

III. HASIL DAN PEMBAHASAN

3.1. Statistik Deskriptif

Analisis deskriptif dalam penelitian ini dilakukan untuk memberikan gambaran umum mengenai data yang digunakan. Hasil ringkasan statistik deskriptif disajikan pada tabel 3.

Tabel 3. Statistik Deskriptif

Variabel	Min	Max	Mean	Median	Standar Deviasi
Y	0	1,317	94.970	45	220.788
X_1	2.380	47.090	10.659	9.360	7.583
X_2	66.19	88.61	75.15	74.02	4.987
X_3	382.4	22,061	3,246	1,134	4,407.350

Berdasarkan tabel 3, jumlah kasus AIDS (Y) menunjukkan variasi yang sangat besar di antara wilayah-wilayah di Pulau Jawa. Jakarta Selatan memiliki jumlah kasus tertinggi (1,317 kasus), diikuti oleh Jakarta Pusat (1,248 kasus) dan Jakarta Timur (1,044 kasus), yang semuanya merupakan wilayah dengan kepadatan penduduk tinggi. Sebaliknya, beberapa wilayah seperti Kepulauan Seribu tidak mencatat kasus AIDS sama sekali, sementara Magelang dan Batu hanya memiliki masing-masing 2 dan 1 kasus, yang mencerminkan perbedaan distribusi penyakit ini antarwilayah.

Variabel sosial ekonomi menunjukkan perbedaan antarwilayah di Pulau Jawa. Persentase penduduk miskin (X1) rata-rata 10,659%, dengan rentang 2,38%–47,09%, mencerminkan kesenjangan kesejahteraan. Daerah seperti Kulon Progo menghadapi tantangan ekonomi lebih besar dibanding wilayah dengan kemiskinan rendah. Indeks Pembangunan Manusia (X2) berkisar antara 66,19 hingga 88,61 (rata-rata 75,15), menunjukkan disparitas pembangunan manusia, dengan Kota Yogyakarta tertinggi dan Sampang terendah. Kepadatan penduduk (X3) sangat bervariasi, dari 382,4 hingga 22.061 jiwa/km², mencerminkan kontras antara wilayah urban dan rural.

3.2. Normalisasi Data

Sebelum masuk ke tahap analisis, data perlu dinormalisasi agar setiap variabel berada dalam skala yang sama. Data dalam penelitian ini dinormalisasi menggunakan metode *Min-Max Normalization*. Hasil normalisasi ditampilkan pada tabel 4.

Tabel 4. Hasil normalisasi data

Y	X_1	X_2	X_3
0.04328	0.19257	0.26093	0.02221
0.06682	0.22702	0.34566	0.04297
0.04176	0.28204	0.19269	0.04117
0.04860	0.28003	0.13247	0.02457
0.04328	0.31223	0.25379	0.03070
0.02733	0.20018	0.36396	0.01599
:	:	:	:
0.08580	0.09595	0.54282	0.42584
:	:	:	:
0.00076	0.02080	0.57449	0.03467

3.3. Regresi Nonparametrik Kernel Fungsi Triangle dengan Estimator Priestley-Chao

3.3.1. Pemilihan *Bandwith* Optimum

Pemilihan bandwidth optimum dalam penelitian ini dilakukan menggunakan metode *Generalized Cross-Validation* (GCV) untuk menentukan nilai *bandwidth* (h) yang memberikan hasil terbaik dalam estimasi regresi nonparametrik kernel. Perhitungan GCV untuk berbagai kombinasi *bandwidth* dilakukan dengan bantuan perangkat lunak Rstudio. Variasi *bandwidth* yang dicoba yaitu dalam rentang 0.1 hingga 3 dengan kenaikan sebesar 0.01. Nilai *bandwidth* yang menghasilkan GCV minimum dipilih sebagai *bandwidth* optimum dan digunakan untuk mengestimasi fungsi regresi nonparametrik kernel dengan fungsi *triangle*.

E-ISSN 2808-5841
P-ISSN 2808-7283

Tabel 5. Hasil perhitungan GCV kernel *triangle*

Kombinasi	h_1	h_2	h_3	GCV
1	3.00	3.00	3.00	0.029582
2	2.50	2.50	3.00	0.029531
3	1.25	1.00	0.50	0.021561
4	1.00	2.50	0.30	0.011689
5	0.40	0.20	0.20	0.011219
6	0.45	0.30	0.15	0.006234
7	0.30	0.15	0.40	0.012579
8	**0.10**	**0.20**	**0.25**	**0.003178**
9	0.10	0.15	0.25	0.003314
⋮	⋮	⋮	⋮	⋮
24,642,171	0.10	0.10	0.10	0.024531

Berdasarkan hasil perhitungan yang ditampilkan pada tabel 5, menunjukkan bahwa nilai GCV minimum yang diperoleh pada kernel *triangle* adalah 0.003178 pada kombinasi *bandwidth* h_1 = 0.10, h_2 = 0.20, dan h_3 = 0.25. Nilai GCV yang lebih kecil menunjukkan bahwa kombinasi *bandwidth* tersebut memberikan estimasi regresi yang lebih baik dibandingkan kombinasi lainnya, karena menghasilkan tingkat kesalahan prediksi yang lebih rendah. Oleh karena itu, kombinasi *bandwidth* ini dipilih sebagai *bandwidth* optimum untuk fungsi *triangle*, yang selanjutnya digunakan dalam proses estimasi model regresi.

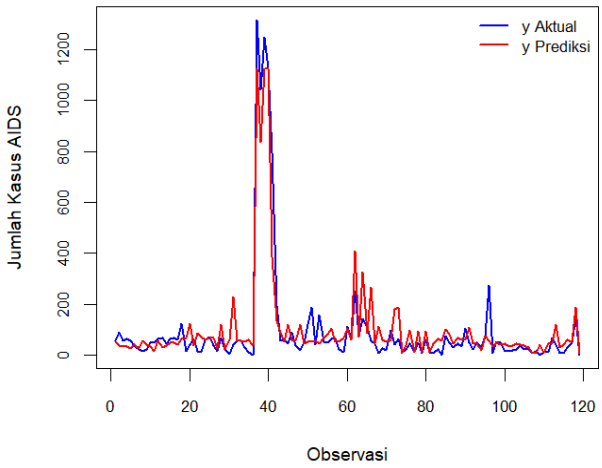
3.3.2. Estimasi Kasus AIDS dengan Fungsi *Triangle*

Menggunakan *bandwidth* optimum, diperoleh persamaan estimasi regresi nonparametrik kernel *Priestley-Chao* dengan fungsi kernel *triangle* sebagai berikut:

$$\begin{aligned}\hat{m}(x) &= \sum_{i=1}^{n}\prod_{j=1}^{m}\left((x_i - x_{i-1})K_h(x - x_i)\right)y_i \\ &= \sum_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{x_{ji}-x_{j(i-1)}}{h_j}\left(\left(1-\left|\frac{x_j-x_{ji}}{h_j}\right|\right)\mathrm{I}\right)\right)y_i \\ &= \sum_{i=1}^{n}\begin{bmatrix}\left(\frac{x_{1i}-x_{1(i-1)}}{0.10}\left(\left(1-\left|\frac{x_1-x_{1i}}{0.10}\right|\right)\mathrm{I}\right)\right)\\ \left(\frac{x_{2i}-x_{2(i-1)}}{0.20}\left(\left(1-\left|\frac{x_2-x_{2i}}{0.20}\right|\right)\mathrm{I}\right)\right)\\ \left(\frac{x_i-x_{3(i-1)}}{0.25}\left(\left(1-\left|\frac{x_3-x_{3i}}{0.25}\right|\right)\mathrm{I}\right)\right)\end{bmatrix}y_i\end{aligned}$$

Diperoleh hasil estimasi yang divisualisasikan pada gambar 2. Grafik tersebut menampilkan perbandingan antara nilai y aktual dan y prediksi untuk melihat sejauh mana model dapat menangkap pola data.

E-ISSN 2808-5841
P-ISSN 2808-7283



Gambar 2. Grafik perbandingan y aktual dan y prediksi (triangle)

Berdasarkan grafik yang ditunjukkan pada gambar 2 dapat dilihat bahwa secara umum, pola y prediksi (garis merah) mengikuti tren y aktual (garis biru), meskipun terdapat beberapa perbedaan. Pada sebagian besar observasi, model mampu menangkap pola utama data dengan baik, khususnya pada saat fluktuasi nilai relatif rendah. Namun, pada sekitar observasi ke-40, terjadi lonjakan pada data yang direspons oleh nilai prediksi, meskipun terdapat perbedaan dalam besarnya perubahan. Hal tersebut menunjukkan bahwa model cukup sensitif terhadap perubahan ekstrem meskipun belum sepenuhnya akurat.

Selain itu, pada observasi dengan nilai rendah, prediksi cenderung stabil dan mendekati nilai y aktual, meskipun terdapat selisih antara keduanya. Secara keseluruhan, model regresi nonparametrik kernel triangle dengan estimator *Priestley-Chao* cukup baik dalam menangkap tren utama, tetapi masih dapat ditingkatkan untuk memodelkan variasi secara lebih akurat.

3.4. Regresi Nonparametrik Kernel Fungsi Gaussian dengan Estimator Priestley-Chao

3.4.1. Pemilihan *Bandwith* Optimum

Setelah melakukan estimasi dengan fungsi *triangle*, langkah berikutnya adalah mencari bandwidth optimum menggunakan fungsi *gaussian*. Hasil perhitungan GCV untuk berbagai kombinasi bandwidth pada fungsi *gaussian* disajikan pada tabel 3.

Tabel 6. Hasil Perhitungan GCV Kernel Gaussian

Kombinasi	h_1	h_2	h_3	GCV
1	3.00	3.00	3.00	0.030015
2	3.00	2.00	2.50	0.030131
3	2.00	1.50	2.50	0.030159
4	1.25	2.00	0.50	0.029316
5	0.50	0.50	0.50	0.027175
6	0.45	0.11	0.10	0.011760
7	0.50	0.30	0.25	0.023480
8	**0.10**	**0.14**	**0.10**	**0.008402**
9	0.40	0.30	0.10	0.010133
⋮	⋮	⋮	⋮	⋮

E-ISSN 2808-5841
P-ISSN 2808-7283

24,642,171	0.10	0.10	0.10	0.008646

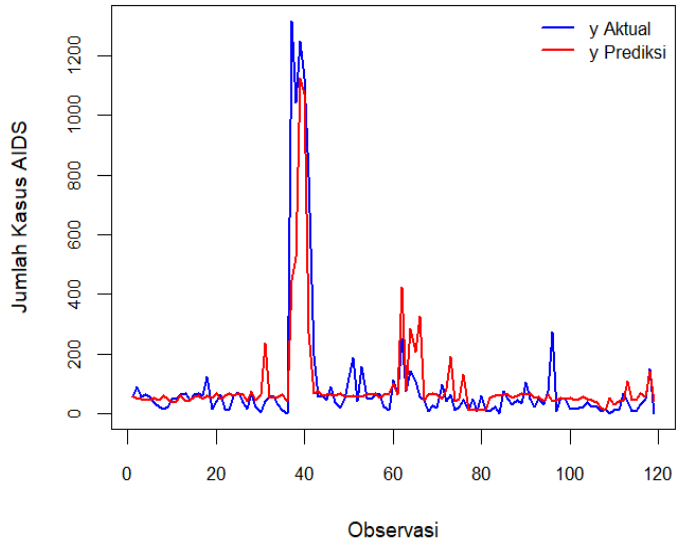
Berdasarkan hasil perhitungan GCV sebagaimana ditunjukkan pada tabel 6 diperoleh nilai minimumnya adalah 0.008402 yaitu pada kombinasi *bandwidth* $h_1 = 0.10$, $h_2 = 0.14$, dan $h_3 = 0.10$. Kombinasi *bandwidth* tersebut digunakan untuk memperoleh persamaan regresi untuk masing-masing fungsi kernel.

3.4.2. Estimasi Kasus AIDS dengan Fungsi *Gaussian*

Diperoleh persamaan estimasi regresi nonparametrik kernel *Priestley-Chao* dengan fungsi *gaussian* sebagai berikut:

$$\begin{aligned}\hat{m}(x) &= \sum_{i=1}^{n}\prod_{j=1}^{m}\left((x_i - x_{i-1})K_h(x - x_i)\right)y_i \\ &= \sum_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{x_{ji}-x_{j(i-1)}}{h_j}\left(\left(1 - \left|\frac{x_j - x_{ji}}{h_j}\right|\right)\mathrm{I}\right)\right)y_i \\ &= \sum_{i=1}^{n}\begin{bmatrix}\left(\frac{x_{1i}-x_{1(i-1)}}{0.10}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_1 - x_{1i}}{0.10}\right)^2\right)\right)\right) \\ \left(\frac{x_{2i}-x_{2(i-1)}}{0.14}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_2 - x_{2i}}{0.14}\right)^2\right)\right)\right) \\ \left(\frac{x_{3i}-x_{3(i-1)}}{0.10}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_3 - x_{3i}}{0.10}\right)^2\right)\right)\right)\end{bmatrix}y_i\end{aligned}$$

Hasil estimasi ditampilkan pada gambar 3. Grafik tersebut memperlihatkan perbandingan antara nilai y aktual dan nilai y prediksi untuk mengevaluasi sejauh mana model mampu mengikuti pola dalam data.



Gambar 3. Grafik perbandingan y aktual dan y prediksi (*gaussian)*

Secara umum terlihat pada gambar 3, model mampu mengikuti pola dari y aktual, terutama pada daerah dengan fluktuasi yang relatif rendah. Pada beberapa observasi, terlihat bahwa prediksi berhasil menangkap tren kenaikan dan penurunan yang terjadi pada data aktual. Namun, model

belum sepenuhnya mampu menangkap nilai puncak yang ekstrem, seperti yang tampak pada observasi sekitar ke-40 dan ke-60, di mana jumlah kasus AIDS aktual mencapai nilai yang sangat tinggi, sementara prediksi cenderung lebih rendah.

Jika dibandingkan dengan hasil estimasi menggunakan fungsi kernel *triangle*, model dengan kernel *gaussian* menunjukkan pola yang lebih halus, terutama pada bagian data dengan fluktuasi tinggi. Pada beberapa observasi, selisih antara nilai prediksi dan aktual masih terlihat cukup jauh, yang mengindikasikan bahwa model belum sepenuhnya menangkap variasi data pada seluruh titik dengan baik.

3.5. Evaluasi Kebaikan Model

Evaluasi *kebaikan* model dilakukan dengan membandingkan nilai *Mean Square Error* (MSE) dan koefisien determinasi (R^2) untuk fungsi kernel *triangle* dan *gaussian*.

Tabel 7. Hasil Perhitungan R^2 dan MSE

Kriteria	*Triangle*	*Gaussian*
Mean Square Error (MSE)	0.003129	0.008247
Koefisien Determinasi (R^2)	0.887730	0.704080

Berdasarkan hasil perhitungan pada tabel 7, model dengan fungsi kernel *triangle* memiliki nilai MSE sebesar 0.003129, sedangkan model dengan fungsi kernel *gaussian* memiliki nilai MSE sebesar 0.008247. Nilai MSE yang lebih kecil pada fungsi kernel *triangle* menunjukkan bahwa model ini menghasilkan tingkat kesalahan prediksi yang lebih rendah dibandingkan dengan fungsi kernel *gaussian*.

Selain itu, nilai koefisien determinasi (R^2) untuk model dengan fungsi kernel *triangle* adalah 0.887730, sementara untuk model dengan fungsi kernel *gaussian* adalah 0.704080. Berdasarkan kategori nilai R^2 seperti pada tabel 2, model dengan fungsi kernel *triangle* masuk dalam kategori akurat ($0.8 \leq R^2 < 0.9$), sedangkan model dengan fungsi kernel *gaussian* masuk dalam kategori cukup akurat ($0.5 \leq R^2 < 0.8$). Hal ini menunjukkan bahwa model dengan kernel *triangle* memiliki kemampuan prediksi yang lebih baik dibandingkan dengan model kernel *gaussian*, karena lebih banyak variasi dalam data yang dapat dijelaskan oleh model. Nilai koefisien determinasi menunjukkan proporsi total variasi pada variabel dependen yang dapat dijelaskan oleh variabel-variabel independen dalam model. Semakin tinggi nilai R^2, semakin besar kontribusi variabel-variabel penjelas dalam menggambarkan pola atau perubahan pada yyy, sehingga model dianggap lebih informatif dan memiliki tingkat kecocokan yang lebih baik [26].

Dengan demikian, berdasarkan kedua kriteria evaluasi yang telah dianalisis, yaitu nilai koefisien determinasi (R^2) dan *Mean Squared Error* (MSE), dapat disimpulkan bahwa model regresi nonparametrik kernel dengan fungsi *triangle* memiliki kinerja yang lebih baik dibandingkan model dengan fungsi kernel *gaussian* dalam menduga hubungan antara variabel.

IV. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan mengenai regresi nonparametrik kernel *triangle* dan *gaussian* dengan estimator *Priestley-Chao* dalam pemodelan kasus AIDS di Pulau Jawa diperoleh kesimpulan sebagai berikut:

1. Nilai GCV minimum yang diperoleh pada model kernel triangle yaitu sebesar 0.003178 pada kombinasi bandwidth h1 = 0.10, h2 = 0.20, dan h3 = 0.25. Sehingga model yang didapatkan adalah:

$$\hat{m}(x) = \sum_{i=1}^{n} \left[\left(\frac{x_{1i} - x_{1(i-1)}}{0.10} \left(\left(1 - \left| \frac{x_1 - x_{1i}}{0.10} \right| \right) \mathrm{I} \right) \right) \left(\frac{x_{2i} - x_{2(i-1)}}{0.20} \left(\left(1 - \left| \frac{x_2 - x_{2i}}{0.20} \right| \right) \mathrm{I} \right) \right) \left(\frac{x_{2i} - x_{3(i-1)}}{0.25} \left(\left(1 - \left| \frac{x_3 - x_{3i}}{0.25} \right| \right) \mathrm{I} \right) \right) \right] y_i$$

2. Nilai GCV minimum yang diperoleh pada model kernel gaussian yaitu sebesar 0.008402 pada kombinasi *bandwidth* h1 = 0.10, h2 = 0.14, dan h3 = 0.10. Sehingga model yang didapatkan adalah:

$$\hat{m}(x) = \sum_{i=1}^{n} \left[\left(\frac{x_{1i} - x_{1(i-1)}}{0.10} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_1 - x_{1i}}{0.10} \right)^2 \right) \right) \right) \left(\frac{x_{2i} - x_{2(i-1)}}{0.14} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_2 - x_{2i}}{0.14} \right)^2 \right) \right) \right) \left(\frac{x_{3i} - x_{3(i-1)}}{0.10} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_3 - x_{3i}}{0.10} \right)^2 \right) \right) \right) \right] y_i$$

3. Berdasarkan nilai *Mean Square Error* (MSE) dan koefisien determinasi (R2), kernel *triangle* menunjukkan kinerja lebih baik dibandingkan kernel *gaussian*. Kernel triangle memiliki MSE sebesar 0.003129, lebih kecil daripada kernel gaussian yang memiliki MSE 0.008247. Selain itu, nilai R2 pada kernel *triangle* sebesar 0.887730, lebih tinggi daripada kernel *gaussian* yang memiliki nilai 0.704080, sehingga menghasilkan estimasi yang mampu menjelaskan variabilitas data dengan lebih baik.
4. Hasil penelitian menunjukkan perbedaan performa antara kernel *triangle* dan *gaussian*. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan kernel lain, seperti *epanechnikov* atau *quartic*, untuk melihat apakah terdapat peningkatan kinerja estimasi.
5. Mengingat bahwa pemilihan bandwidth berpengaruh terhadap kualitas estimasi, disarankan untuk mengeksplorasi metode lain dalam menentukan bandwidth yang lebih optimal guna meningkatkan akurasi model, seperti *Least Square Cross Validation*.

REFERENSI

1. Subowo, *Immunologi Klinik*, 2nd ed. Jakarta: SagungbSeto, 2013.
2. N. Agustina and Kemenkes RI, "Direktorat Jenderal Pelayanan Kesehatan," 2022. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/1777/stress-dan-penyebabnya%0Ahttps://yankes.kemkes.go.id/view_artikel/1767/stop-tuberkulosis%0Ahttps://yankes.kemkes.go.id/view_artikel/1222/gula-si-manis-yang-menyebabkan-ketergantungan%0Ahttps://yankes.kemkes.go
3. WHO, "HIV and AIDS," World Health Organization (WHO). [Online]. Available: https://www-who-int.translate.goog/news-room/fact-sheets/detail/hiv-aids?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
4. N. Afriana *et al.*, *Laporan Tahunan HIV AIDS 2022*. 2023. [Online]. Available: http://p2p.kemkes.go.id/wp-content/uploads/2023/06/FINAL_6072023_Layout_HIVAIDS-1.pdf

5. Kemenkes RI, "Laporan Eksekutif Perkembangan HIV AIDS dan Penyakit Infeksi Menular Seksual (PIMS) Triwulan I Tahun 2023," Sistem Informasi HIV AIDS, 2023.
6. Badan Pusat Statistik, *Statistik Indonesia 2024*, vol. 52. 2024. [Online]. Available: https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html
7. N. Muhamad, "Ada 16 Ribu Kasus AIDS Baru di Indonesia, Terbanyak di Jawa Barat," 2024.
8. C. L. Soo, N. P. Pai, S. J. Bartlett, A. Esmail, K. Dheda, and S. Bhatnagar, "Socioeconomic Factors Impact the Risk of HIV Acquisition in the Township Population of South Africa: A Bayesian Analysis," *PLOS Global Public Health*, vol. 3, no. 1, pp. 1–9, 2023, doi: 10.1371/journal.pgph.0001502.
9. N. S. Khairunisa and E. D. Sihaloho, "Determinan Pembangunan Daerah dan Angka HIV/AIDS di Indonesia," *Ekonomikawan: Jurnal Ilmu Ekonomi dan Studi Pembangunan*, vol. 19, no. 1, pp. 42–56, 2019, doi: 10.30596/ekonomikawan.v19i1.3084.
10. A. Azizah and I. N. Budiantara, "Faktor-faktor yang Mempengaruhi Persentase Kasus HIV di Indonesia Tahun 2017 Menggunakan Regresi Nonparametrik Spline Truncated," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.44382.
11. S. F. Thahar and T. Sirait, "Analisis Spasial Variabel-Variabel yang Memengaruhi Jumlah Kasus Baru HIV/AIDS di Provinsi Jawa Timur Tahun 2021," *Seminar Nasional Official Statistics*, vol. 2023, no. 1, pp. 287–296, 2023, doi: 10.34123/semnasoffstat.v2023i1.1602.
12. J. S. Armstrong, "Illusions in regression analysis," *Int J Forecast*, vol. 28, no. 3, pp. 689–694, 2012, doi: https://doi.org/10.1016/j.ijforecast.2012.02.001.
13. E. D. A. A. Sita, "Pendekatan Multivariate Adaptive Regression Splines (Mars) Pada Data Panel Untuk Pemodelan Penduduk Miskin Di Indonesia," vol. 2012, no. November, pp. 175–191, 2014.
14. R. L. Eubank, *Spline smoothing and Nonparametrik Regression*. New York: Marcel Dekker, 1988.
15. W. Härdle, *Applied Nonparametric Regression*. in Econometric Society Monographs. Cambridge University Press, 1990.
16. S. Wulandary and D. I. Purnama, "Perbandingan Regresi Nonparametrik Kernel Dan B-Splines Pada Pemodelan Rata-Rata Lama Sekolah Dan Pengeluaran Perkapita Di Indonesia," *Jambura Journal of Probability and Statistics*, vol. 1, no. 2, pp. 89–97, 2020, doi: 10.34312/jjps.v1i2.7501.
17. W. Härdle, *Smoothing Techniques*. in Springer Series in Statistics. New York, NY: Springer New York, 1991. doi: 10.1007/978-1-4612-4432-5.
18. M. Rizaldi, "Estimasi Kurva Regresi Nonparametrik Fungsi Kernel Pada Data Longitudinal (Studi Kasus : Laju Pertumbuhan Ekonomi Provinsi Nusa Tenggara Barat 2016-2020) Estimasi Kurva Regresi Nonparametrik Fungsi Kernel Pada Data Longitudinal (Studi Kasus : Laju Pert," 2023.
19. S. Nisa', "Estimator Kernel Epanechnikov Dan Kernel Triangle Pada Data Rata-Rata Bulanan Bilangan Sunspot, NOAA," Universitas Islam Negeri Maulana Malik Ibraahim Malang, 2016.
20. R. Handayani, "Analisis Regresi Kernel dengan Estimator Nadaraya-Watson," 2019.
21. A. T. Ampa, I. Monica, Makkulau, L. O. Saidi, and N. Muhtar, "Priestley Chao Estimator in Nonparametric Multivariable Kernel Regression in Estimating The Value of Indonesia's Balance Trade," *ITM Web of Conferences*, vol. 58, p. 04002, 2024, doi: 10.1051/itmconf/20245804002.
22. T. A. Salsabela, "Regresi nonparametrik kernel menggunakan penaksir Priestley-Chao untuk memodelkan inflasi di Indonesia," 2023, [Online]. Available: http://etheses.uin-malang.ac.id/51112/%0Ahttp://etheses.uin-malang.ac.id/51112/1/19610101.pdf
23. I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K. K. Singh, "A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain," *Mathematics*, vol. 10, no. 11, 2022, doi: 10.3390/math10111942.
24. G. Fudllayati, "Analisis Produk Domestik Bruto (PDB) Dengan Regresi Nonparametrik Kernel Menggunakan Estimator Priestley-Chao," 2021.
25. S. F. Sayuti, "Regresi Nonparametrik Metode Kernel Dengan Estimator Gasser-Muller Pada Data Curah Hujan Di Provinsi Lampung," 2022.
26. Shalabh and S. S. Dhar, "Goodness of Fit in Nonparametric Regression Modelling," *J Stat Theory Pract*, vol. 15, no. 1, pp. 1–13, 2021, doi: 10.1007/s42519-020-00148-x.