

PEMILIHAN PARAMETER OSILASI OPTIMAL MENGGUNAKAN *GENERALIZED CROSS-VALIDATION (GCV) PADA REGRESI NONPARAMETRIK DERET FOURIER*

Hasna Faridah Dhiya Ul Haq^{1*}, I Nyoman Budiantara², dan Jerry Dwi Trijoyo Purnomo³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember

*e-mail: hasnafrd01@gmail.com

DOI: [10.14710/j.gauss.14.2.577-587](https://doi.org/10.14710/j.gauss.14.2.577-587)

Article Info:

Received: 2025-10-21

Accepted: 2025-12-10

Available Online: 2025-12-13

Keywords:

Fourier Series, Generalized Cross-Validation (GCV), Nonparametric Regression, Poverty

Abstract: This research focuses on determining the optimal oscillation parameter in a Fourier series nonparametric regression model using the Generalized Cross-Validation (GCV) method to analyze factors influencing poverty in Yogyakarta, Central Java, and West Java Provinces in 2024. The dataset indicates a correlation between predictor and response variables that exhibits a recurring pattern and does not follow a particular distribution, making Fourier series suitable because it can flexibly approximate smooth yet complex functional relationships without imposing strict parameter assumptions. GCV is used to identify optimal oscillation, to achieve a balance between model fit and complexity level. The results show that the minimum GCV is achieved at two oscillations, producing the minimum GCV value of 7.617 and an MSE of 5.708, reflecting good predictive performance. Simultaneous testing shows a significant joint effect of predictors, while partial testing reveals that only some Fourier components are statistically significant. Residual diagnostic tests show that the assumptions of homoscedasticity, independence, and normality are satisfied, indicating that the fitted model is statistically reliable. This study demonstrates that the Fourier series approach with GCV optimization provides an effective alternative to conventional parametric methods for capturing nonlinear relationships in regional poverty data with minimal distributional assumptions.

1. PENDAHULUAN

Analisis regresi adalah metode yang digunakan untuk mengenali pola hubungan antara sejumlah variabel dengan tujuan untuk menentukan bentuk estimasi kurva regresi pada statistika. Regresi parametrik, regresi nonparametrik, dan regresi semiparametrik adalah beberapa metode yang telah dikembangkan oleh peneliti berdasarkan bentuk kurva regresinya (Fernandes dkk., 2021). Dalam banyak kasus nyata, keterkaitan antara variabel respon dengan variabel prediktor tidak menunjukkan pola spesifik atau cenderung berbentuk acak, sehingga pendekatan yang tepat untuk dipakai yaitu regresi nonparametrik. Pemodelan regresi nonparametrik memiliki beberapa estimator yang sering digunakan yakni Deret *Fourier*, *Kernel*, dan *Spline*. Salah satu estimator yang banyak digunakan yaitu deret *Fourier*.

Deret *Fourier* baik digunakan apabila pola data memiliki kecenderungan berulang. Estimator deret *Fourier* ini membangun model regresi dengan menggunakan fungsi-fungsi sinus dan cosinus sebagai basis fungsi. Pada deret *Fourier* sendiri, terdapat tantangan pada penggunaannya, yaitu pada penentuan parameter osilasi yang hasilnya akan menentukan bentuk kurva yang dihasilkan. Jika jumlah osilasi yang digunakan masih belum menyentuh titik optimal, maka model akan menjadi terlalu kaku dan tidak dapat menangkap pola data dengan baik (*underfitting*). Sebaliknya, jika jumlah osilasi melebihi titik optimal, model

justru akan terlalu menyesuaikan diri dengan data hingga mengikuti *noise (overfitting)*. Dengan demikian, penentuan jumlah parameter osilasi yang paling optimal menjadi hal yang krusial untuk mendapat model yang akurat.

Sejumlah metode yang dapat digunakan untuk menentukan jumlah osilasi yaitu *Cross-Validation* (CV), *Unbiased Risk* (UBR), dan *Generalized Cross-Validation* (GCV). GCV merupakan metode yang meminimumkan jumlah kuadrat galat prediksi dengan memberikan bobot yang sama untuk tiap pengamatan, sehingga mampu menyeimbangkan kompleksitas model dan tingkat kesalahan pada estimasi. Metode GCV sering dipakai karena merupakan metode yang lebih stabil dibandingkan CV karena sejatinya, metode GCV adalah bentuk pengembangan dari metode CV (Wahba, 1990), serta lebih praktis jika dibandingkan dengan metode UBR.

Penelitian sebelumnya telah menyoroti mengenai pemilihan parameter optimal dengan menggunakan GCV. Sebagaimana penelitian oleh Utami dkk (2019) yang menunjukkan bahwa metode GCV lebih unggul dibandingkan UBR dalam menetapkan titik knot paling optimal dengan regresi nonparametrik *Spline*. Sementara itu, pada penelitian Amrullah & Amalia (2022) yang meneliti perbandingan GCV dengan CV untuk data kemiskinan di Provinsi Sumatra Utara dengan regresi nonparametrik deret *Fourier* menghasilkan bahwa metode GCV lebih baik dibandingkan CV. Penelitian Ni'matuzzahroh & Dani (2024) menggunakan GCV dengan regresi nonparametrik deret *Fourier* menggunakan data Pendidikan menunjukkan bahwa deret *Fourier* mampu diaplikasikan pada permasalahan di Indonesia. Meskipun demikian, pemanfaatan deret *Fourier* pada isu sosial-ekonomi masih terbatas, sehingga penerapan model ini pada konteks berbeda tetap diperlukan untuk memperkuat bukti empiris.

Kemiskinan masih menjadi isu yang penting di seluruh negara di dunia. Kemiskinan adalah salah satu hal yang penting untuk dipedulikan karena jumlahnya yang masih tinggi di Indonesia. Pada wilayah Pulau Jawa, tiga provinsi dengan tingkat kemiskinan tertinggi pada tahun 2024 yaitu Daerah Istimewa Yogyakarta (DIY), Jawa Tengah, dan Jawa Barat. Ketiganya menunjukkan persentase penduduk miskin yang relatif tinggi yaitu sebesar 10,83%, 10,47%, dan 9,79% yang mana angka-angka tersebut masih tergolong tinggi, sehingga analisis mendalam dibutuhkan untuk memahami faktor-faktor yang mempengaruhi kemiskinan. Kemiskinan sendiri dianalisis berdasarkan beberapa indikator, seperti ekonomi, ketenagakerjaan, dan pendidikan untuk memahami hubungan antarvariabel secara fleksibel melalui pendekatan deret *Fourier*.

Penerapan pendekatan dengan menggunakan regresi nonparameter deret *Fourier* pada isu kemiskinan pada ketiga provinsi tersebut masih banyak belum dieksplorasi. Pada kondisi tahun 2024, hubungan antara persentase penduduk miskin dengan indikator-indikator yang diduga sebagai faktor penyebabnya menunjukkan karakteristik hubungan yang tidak menunjukkan pola, sehingga kurang sesuai jika didekati dengan model parametrik. Kondisi ini mengakibatkan pengaplikasian regresi nonparametrik deret *Fourier* cocok digunakan karena mampu menangkap pola yang tidak mengikuti distribusi serta cenderung berulang.

Penelitian ini bertujuan untuk menentukan jumlah osilasi optimal pada model regresi nonparametrik deret *Fourier* sehingga dapat memberikan hasil terbaik dalam memprediksi tingkat kemiskinan di ketiga provinsi tersebut. Hasil penelitian ini diharapkan bisa berkontribusi pada pengembangan metode pemilihan parameter optimal pada regresi nonparametrik serta memberikan manfaat bagi perumusan kebijakan pengentasan kemiskinan.

2. TINJAUAN PUSTAKA

Kemiskinan merupakan keadaan saat individu atau sekumpulan orang tidak mempunyai kapabilitas yang cukup untuk memenuhi keperluan dasar yang minimum, baik keperluan pangan maupun non-pangan guna mencapai kehidupan yang layak (BPS, 2011). Pengertian kemiskinan terus berkembang bersamaan dengan semakin kompleksnya aspek yang mempengaruhi dan permasalahan lain yang menyertainya. Pada saat ini, kemiskinan tidak lagi dipandang semata-mata sebagai persoalan ekonomi, tetapi juga mencangkup aspek sosial, kesehatan, politik, dan pendidikan. Pada penelitian Saputro & Arif (2024) menghasilkan bidang pendidikan dan bidang ketenagakerjaan berperan terhadap kemiskinan. Serta Pynanjung, dkk (2021) menjelaskan bahwa bidang ekonomi dan bidang ketenagakerjaan berpengaruh terhadap kemiskinan.

Bidang ekonomi merupakan bidang yang dianggap sebagai salah satu pilar pembangunan berkelanjutan yang dapat digunakan untuk mengatasi kemiskinan. Untuk mengukur aspek pembangunan berkelanjutan ini menggunakan PDRB (Produk Domestik Regional Bruto) (Laswinia & Chamid, 2016). Penelitian Sari & Novianti (2024) melakukan penelitian mengenai pengaruh PDRB terhadap kemiskinan yang menghasilkan PDRB berpengaruh secara negatif terhadap kemiskinan.

Pendidikan berperan penting dalam upaya untuk mengurangi tingkat kemiskinan dalam jangka panjang. Upaya tersebut dapat dilakukan secara tidak langsung melalui peningkatan produktivitas, maupun secara langsung dengan memberikan pelatihan keterampilan yang relevan bagi kelompok masyarakat miskin agar kemampuan dan pendapatan mereka meningkat. Pada berbagai penelitian, Rata-Rata Lama Sekolah (RLS) merupakan tingkat pendidikan yang diraih masyarakat di suatu daerah. Pernyataan ini didukung oleh penelitian Saputro & Arif (2024) yang meneliti mengenai pengaruh RLS terhadap persentase penduduk miskin yang menghasilkan bahwa RLS berpengaruh signifikan terhadap persentase penduduk miskin.

Ketenagakerjaan merupakan bidang lain yang diprediksi berpengaruh terhadap kemiskinan. Bidang ini direpresentasikan oleh Tingkat Pengangguran Terbuka (TPT). TPT yaitu persentase dari jumlah pengangguran terhadap total angkatan kerja. Pada penelitian Suganda, dkk (2024) yang meneliti mengenai kemiskinan menghasilkan bahwa TPT berpengaruh signifikan terhadap kemiskinan.

Regresi nonparametrik adalah metode untuk menganalisis bentuk hubungan antar variabel ketika bentuk fungsi regresinya tidak diketahui dan informasi mengenai pola data masa lalu tidak tersedia secara lengkap (Eubank, 1999). Bentuk model dari regresi nonparametrik yaitu:

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

dengan y_i yaitu variabel respon pada pengamatan ke- i , x_i yaitu variabel prediktor pada pengamatan ke- i , $g(x_i)$ yaitu fungsi regresi nonparametrik untuk prediktor pada pengamatan ke- i , dan ε_i yaitu *random error* yang diperkirakan identik, independen, dan berdistribusi normal yang ditulis dengan $\varepsilon_i \sim IIDN(0, \sigma^2)$.

Pada regresi nonparametrik deret *Fourier*, fungsinya dituliskan seperti berikut.

$$g(x_i) = bx_i + \frac{1}{2}a_0 + \sum_{s=1}^S a_s \cos sx_i \quad (2)$$

dengan $b, a_0, a_1, \dots, a_s, s = 1, 2, \dots, S$ merupakan parameter yang didekati dengan fungsi deret *Fourier* serta s adalah banyak parameter osilasi. Pendekatan regresi nonparametrik deret *Fourier* didapatkan dengan meminimalkan *Least Square* (LS).

Model regresi nonparametrik deret *Fourier* tersebut dituliskan pada persamaan (3).

$$y_i = bx_i + \frac{1}{2}a_0 + \sum_{s=1}^S a_s \cos sx_i + \varepsilon_i \quad (3)$$

Model regresi nonparametrik tersebut dijabarkan dalam bentuk persamaan vektor dan matriks seperti pada Persamaan (4).

$$\begin{aligned} \mathbf{y} &= \mathbf{H}(\mathbf{S})\mathbf{a} + \boldsymbol{\varepsilon} & (4) \\ \mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \\ \mathbf{a}' &= [b_1 \quad a_{01} \quad a_{11} \quad a_{21} \quad \cdots \quad a_{S1} \quad | \quad \cdots \quad | \quad b_q \quad a_{0q} \quad a_{1q} \quad \cdots \quad a_{Sq}]_{1 \times q(S+2)} \\ \mathbf{H}(\mathbf{S}) &= \begin{bmatrix} x_{11} & \frac{1}{2} \cos x_{11} & \cdots & \cos Sx_{11} & | & \cdots & | & x_{q1} & \frac{1}{2} \cos x_{q1} & \cdots & \cos Sx_{q1} \\ x_{12} & \frac{1}{2} \cos x_{12} & \cdots & \cos Sx_{12} & | & \cdots & | & x_{q2} & \frac{1}{2} \cos x_{q2} & \cdots & \cos Sx_{q2} \\ \vdots & \vdots & \ddots & \vdots & | & \ddots & | & \vdots & \vdots & \ddots & \vdots \\ x_{1n} & \frac{1}{2} \cos x_{1n} & \cdots & \cos Sx_{1n} & | & \cdots & | & x_{qn} & \frac{1}{2} \cos x_{qn} & \cdots & \cos Sx_{qn} \end{bmatrix}_{n \times q(S+2)} \end{aligned}$$

Selanjutnya parameter \mathbf{a} dapat diestimasi dengan metode *Ordinary Least Square* (OLS) yakni metode estimasi dengan cara meminimalkan jumlah kuadrat *error* dan estimator yang dihasilkan adalah sebagai berikut.

$$\hat{\mathbf{a}} = (\mathbf{H}'(\mathbf{S})\mathbf{H}(\mathbf{S}))^{-1}\mathbf{H}'(\mathbf{S})\mathbf{y} \quad (5)$$

maka,

$$\begin{aligned} \hat{f}(S) &= \mathbf{H}(\mathbf{S})\hat{\mathbf{a}} \\ &= \mathbf{H}(\mathbf{S})(\mathbf{H}'(\mathbf{S})\mathbf{H}(\mathbf{S}))^{-1}\mathbf{H}'(\mathbf{S})\mathbf{y} \\ &= \mathbf{A}(\mathbf{S})\mathbf{y} \end{aligned}$$

$$\mathbf{A}(\mathbf{S}) = \mathbf{H}(\mathbf{S})(\mathbf{H}'(\mathbf{S})\mathbf{H}(\mathbf{S}))^{-1}\mathbf{H}'(\mathbf{S})$$

Pemilihan parameter osilasi optimum pada regresi nonparametrik dengan estimator deret *Fourier* dengan metode *Generalized Cross-Validation* (GCV) yaitu dengan membagi nilai *error* rata-rata dengan proporsi derajat kebebasan residu yang memberikan penyesuaian terhadap kompleksitas model. Pemilihan model terbaik dan optimal didapat dari nilai GCV minimum. Wu dan Zhang (2006) menjadikan persamaan GCV sebagai berikut.

$$GCV = \min_S \left[\frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{[n^{-1} \text{trace}\{\mathbf{I} - \mathbf{A}(\mathbf{S})\}]^2} \right] \quad (5)$$

dimana y_i adalah nilai aktual, \hat{y}_i adalah nilai prediksi model, \mathbf{I} yaitu matriks identitas, $\mathbf{A}(\mathbf{S})$ yaitu matriks untuk menaksir estimator, dan n yaitu banyak pengamatan.

Setelah didapatkan parameter osilasi optimal, maka selanjutnya dapat melakukan pengecekan seberapa besar *error* yang dihasilkan pada model dengan menggunakan MSE. Rumus untuk mendapatkan nilai MSE yaitu pada Persamaan (6).

$$MSE = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

Semakin rendah nilai MSE yang diperoleh suatu model, maka kualitas model tersebut dianggap semakin baik.

Uji hipotesis dilakukan untuk melihat efek yang dihasilkan dari variabel respon terhadap variabel prediktor. Uji hipotesis ini ada dua yaitu uji secara simultan atau pengujian yang dilakukan secara bersamaan dan uji secara parsial yaitu pengujian yang dilakukan pada masing-masing variabel. Pada uji simultan, apabila nilai signifikansi probabilitas $\leq \alpha$, maka

variabel prediktor secara simultan berpengaruh signifikan terhadap variabel respon. Dengan demikian, hipotesis uji simultan dapat dituliskan:

$$H_0 : a_1 = a_2 = \dots = a_n = 0$$

(Variabel prediktor tidak berpengaruh pada variabel respon secara simultan)

$$H_1 : \text{Paling sedikit ada satu } a_i \neq 0; i = 1, \dots, n$$

(Variabel prediktor berpengaruh pada variabel respon secara simultan)

Statistik uji:

$$F_{hitung} = \frac{MSR}{MSE} \quad (7)$$

dengan $MSR = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{p}$ dan $MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1}$ serta tolak H_0 jika $F_{hitung} > F_{tabel(p, n-p-1; \alpha)}$ atau $p-value \leq \alpha$.

Setelah melakukan uji secara simultan, selanjutnya menguji secara parsial atau secara masing-masing. Hipotesis untuk pengujian ini dapat dilihat sebagai berikut:

$$H_0 : a_i = 0$$

(Variabel prediktor tidak berpengaruh pada variabel respon secara parsial)

$$H_1 : a_i \neq 0; i = 1, \dots, n$$

(Variabel prediktor berpengaruh pada variabel respon secara parsial)

Statistik Uji:

$$t_{hitung} = \frac{\hat{a}_i}{SE(\hat{a}_i)} \quad (8)$$

dengan $SE(\hat{a}_i)$ adalah standar *error* \hat{a}_i yang didapatkan dengan Persamaan (9) berikut:

$$SE(\hat{a}_i) = \sqrt{var(\hat{a}_i)} \quad (9)$$

H_0 ditolak jika nilai $|t_{hitung}| > t_{tabel(\frac{\alpha}{2}; n-p-1)}$ atau saat $p-value \leq \alpha$ (Sari & Budiantara, 2012).

Selanjutnya melakukan uji asumsi residual, yang bertujuan untuk menilai apakah model terbaik yang diperoleh memenuhi asumsi identik, independen, serta berdistribusi normal. Tahap pertama adalah uji asumsi identik untuk menentukan homogenitas varians residual dari model regresi. Uji ini menggunakan uji *Glejser* seperti berikut (Octavanny dkk, 2017).

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = \sigma^2 \text{ (Residual identik)}$$

$$H_1 : \text{Minimal terdapat satu } \sigma_i^2 \neq \sigma^2; i = 1, \dots, n \text{ (Residual tidak identik)}$$

Statistik Uji:

$$F_{hitung} = \frac{\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}_i|)^2 / (p-1)}{\sum_{i=1}^n (|e_i| - |\hat{e}_i|)^2 / (n-p)} \quad (10)$$

dengan $|e_i|$ yaitu nilai *error* mutlak. H_0 ditolak jika nilai $F_{hitung} < F_{tabel(p, n-p-1; \alpha)}$ atau $p-value > \alpha$. Lalu, ada uji asumsi independen untuk menentukan apakah ada autokorelasi pada residual. Uji yang digunakan yaitu *run test*, dengan hipotesisnya yaitu:

$$H_0 : \text{Residualnya acak yang artinya tidak ada autokorelasi atau residual independen}$$

$$H_1 : \text{Residualnya tidak acak yang artinya ada autokorelasi atau residual dependen}$$

Statistik uji pada uji ini yaitu r (banyak runtun yang terjadi). H_0 ditolak jika $r > r_{atas}$ atau $r > r_{bawah}$, dimana nilai r_{atas} dan r_{bawah} berasal dari tabel nilai kritis pada runtun r dengan n_1 dan n_2 , yang mana n_1 yaitu jumlah observasi yang nilainya melebihi rata-rata, dan n_2 yaitu jumlah observasi yang nilainya kurang dari rata-rata (Daniel, 1989).

Tahap terakhir ada uji asumsi distribusi normal menggunakan uji *Kolmogorov Smirnov* sebagai berikut (Daniel, 1989).

$$H_0 : F_n(x) = F_0(x) \text{ (Residual berdistribusi normal)}$$

$H_1 : F_n(x) \neq F_0(x)$ (Residual tidak berdistribusi normal)
Statistik uji:

$$D = \sup_x |F_n(x) - F_0(x)| \quad (11)$$

dimana $F_n(x)$ yakni distribusi frekuensi kumulatif dan $F_0(x)$ yakni distribusi frekuensi kumulatif observasi. H_0 ditolak jika $D > D_{(1-\alpha;n)}$ atau $p-value < \alpha$.

3. METODE PENELITIAN

Data yang dipakai adalah data sekunder yang diambil dari publikasi Kemiskinan Tahun 2024 dari Badan Pusat Statistik. Variabel respon yang dipakai yaitu persentase penduduk miskin (Y), sedangkan variabel prediktor (X) yang digunakan yaitu PDRB, RLS, dan TPT.

Langkah untuk mendapatkan hasil dari tujuan penelitian ini yaitu sebagai berikut:

- a. Menganalisis statistik deskriptif pada tiap variabel prediktor serta variabel respon.
- b. Membentuk *scatterplot* antara variabel respon dengan tiap variabel prediktor.
- c. Membangun model dari data persentase penduduk miskin di Provinsi DIY, Jawa Tengah, dan Jawa Barat pada tahun 2024 dengan menggunakan pendekatan nonparametrik estimator deret *Fourier* yang ditunjukkan pada Persamaan (3).
- d. Mendapatkan parameter osilasi yang optimal dengan metode GCV dengan rumus ada di Persamaan (5). Kemudian menetapkan model terbaik dengan melihat nilai GCV terkecil.
- e. Menilai tingkat kesalahan yang dihasilkan model terbaik berdasarkan nilai MSE yang didapatkan dari Persamaan (6).
- f. Membuat grafik perbandingan antara data prediksi dengan data aktual.
- g. Melakukan uji hipotesis berdasarkan model terbaik yang didapatkan secara simultan yang didapatkan dari Persamaan (7) dan parsial yang didapat dari Persamaan (8).
- h. Melakukan uji asumsi residual yaitu identik yang didapat dari Persamaan (10), independen, dan berdistribusi normal yang didapat dari Persamaan (11). Jika asumsi tidak terpenuhi, maka lakukan transformasi data lalu mengulang dari langkah (c).

4. HASIL DAN PEMBAHASAN

Berdasarkan langkah-langkah untuk menganalisis data, langkah pertamanya adalah analisis deskriptif pada tiap variabel penelitian yang dipakai. Langkah ini digunakan untuk eksplorasi data sehingga dapat terlihat gambaran mengenai data. Analisis deskriptif yang digunakan yaitu rata-rata, standar deviasi, nilai maksimum, dan nilai minimum. Hasil tersebut ada pada Tabel 1.

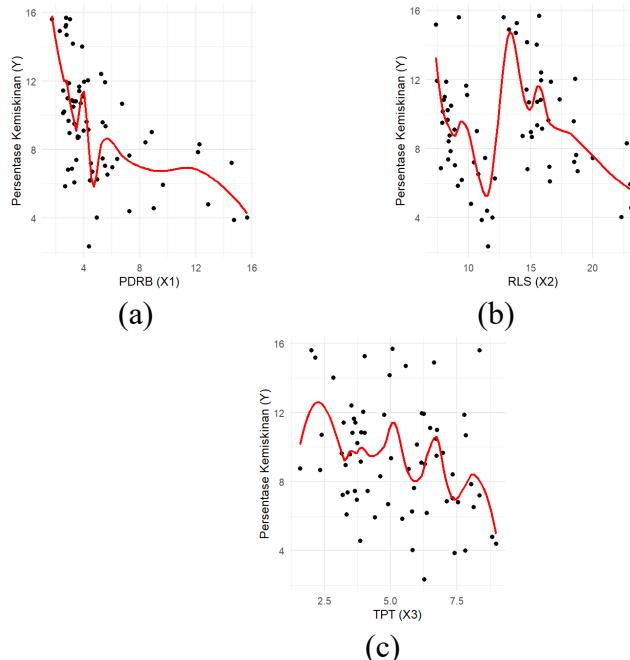
Tabel 1. Analisis Statistik Deskriptif Variabel-Variabel Penelitian

Variabel	Rata-Rata	Standar Deviasi	Minimum	Maksimum
Persentase Penduduk Miskin (Y)	9,36	3,22	2,34	15,71
PDRB (X ₁)	5,11	3,20	1,69	15,66
RLS (X ₂)	13,19	4,30	7,35	23,10
TPT (X ₃)	5,25	1,89	1,58	8,97

Berdasarkan Tabel 1, persentase penduduk miskin di ketiga provinsi tahun 2024 memiliki rata-rata sebesar 9,36% dengan standar deviasinya 3,22%, lalu memiliki nilai persentase penduduk miskin terkecil senilai 2,34% yang berada di Kota Depok, sedangkan persentase penduduk miskin terbesar yaitu 15,71% di Kabupaten Kebumen.

Langkah selanjutnya yaitu melakukan eksplorasi data melalui pembuatan *scatterplot* yang bertujuan untuk menunjukkan pola hubungan antara variabel respon dan setiap variabel

prediktor. Gambar 1 menampilkan hasil *scatterplot* yang merepresentasikan hubungan variabel respon dengan ketiga variabel prediktor tersebut.



Gambar 1. *Scatterplot* Pola Hubungan antara (a) Persentase Penduduk Miskin dan PDRB, (b) Persentase Penduduk Miskin dan RLS, serta (c) Persentase Penduduk Miskin dan TPT

Berdasarkan Gambar 1, terlihat bahwa pola hubungan antara variabel respon dengan variabel-variabel prediktor cenderung berulang yang sesuai dengan kurva deret *Fourier* yang bentuknya cenderung tidak mengikuti pola distribusi tertentu dan berulang. Setelah membuat deskriptif terhadap variabel-variabel penelitian, maka selanjutnya melakukan penentuan parameter osilasi optimal untuk mendapatkan estimator deret *Fourier* paling optimal pada regresi nonparametrik. Berikut merupakan nilai GCV dengan MSE.

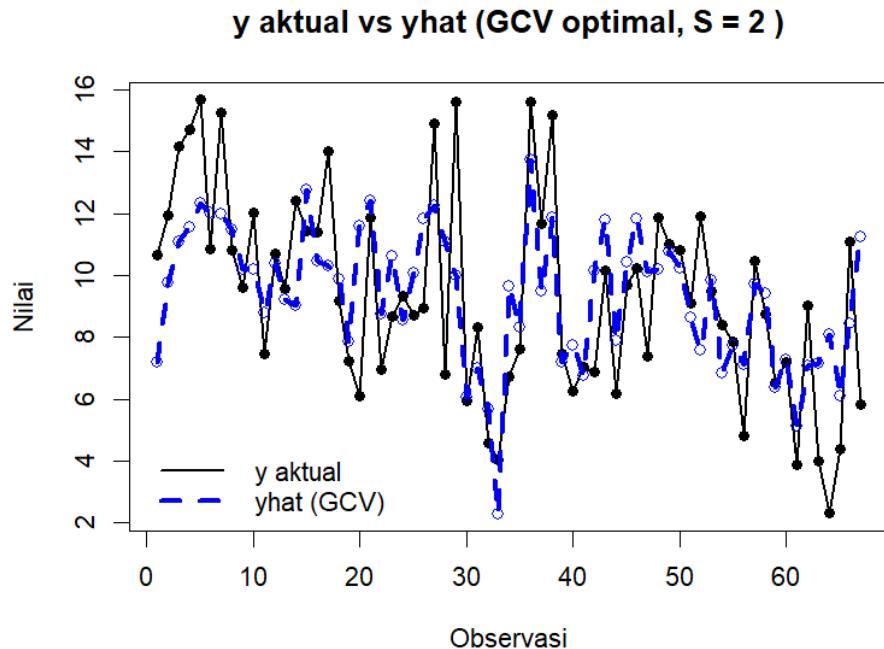
Tabel 2. Hasil Nilai GCV dan MSE

Osilasi (S)	GCV	MSE
1	8,191	6,568
2	7,617	5,708
3	7,784	5,634
4	7,756	5,418
5	7,736	5,405
6	7,785	5,246
7	7,776	5,240
8	7,827	5,084
9	8,087	5,060
10	8,095	5,065

Tabel 2 memperlihatkan bahwa nilai GCV minimum berada pada 2 osilasi dengan nilai GCV sebesar 7,617 dengan nilai MSE sebesar 5,708 yang berarti model yang dihasilkan pada 2 osilasi menghasilkan rata-rata kesalahan prediksi sebesar 5,708 atau sebesar 2,39% dari nilai aktualnya. Maka model yang bisa dituliskan berdasarkan osilasi paling optimal yang didapatkan berdasarkan nilai GCV minimum dapat dituliskan sebagai berikut

$$\hat{y} = -30730,663 - 4,181x_1 + 57156,824 \cos x_1 - 14895,606 \cos 2x_1 - 2,615x_2 + 9993,462 \cos x_2 - 2753,908 \cos 2x_2 + 0,883x_3 - 45774,869 \cos x_3 + 11669,783 \cos 2x_3 \quad (12)$$

Berdasarkan hasil persamaan (12), didapatkan perbandingan antara data prediksi dengan data aktualnya yang divisualisasikan pada Gambar 2.



Gambar 2. Visualisasi Data Prediksi dengan Data Aktual

Berdasarkan visualisasi pada Gambar 2, terlihat pola data yang dihasilkan dari prediksi model \hat{y} yang dipilih dengan menggunakan kriteria GCV optimal dengan 2 osilasi cenderung mengikuti bentuk pola data aktual dengan cukup baik meskipun memiliki perbedaan pada beberapa observasi.

Langkah selanjutnya yaitu menguji hipotesis baik secara simultan maupun parsial terhadap model yang dibentuk. Tabel 3 berikut merupakan hasil uji secara simultan.

Tabel 3. Hasil Uji Simultan

Sumber Variasi	Derajat Kebebasan	Jumlah Kuadrat	Rata-Rata Kuadrat	F_{hitung}	$p-value$
Regresi	3	311,731	103,910	17,118	0,000
Error	63	382,429	6,070		
Total	66	694,161			

Dari Tabel 3 terlihat bahwa nilai F_{hitung} yang didapat sebesar 17,118 yang nilainya lebih besar dari nilai nilai $F_{tabel}(3,63;0,05)$ sebesar 2,75 serta $p - value (0,000) < \alpha(0,05)$, yang artinya H_0 ditolak. Oleh karena itu, dapat diketahui bahwa PDRB, RLS, dan TPT berpengaruh terhadap persentase penduduk miskin secara simultan.

Setelah didapatkan hasil uji simultan, langkah selanjutnya yaitu melakukan uji parsial. Tabel 4 menampilkan hasil dari uji parsial untuk masing-masing variabel prediktor.

Tabel 4. Hasil Pengujian Secara Parsial

Variabel	Parameter	Estimasi	t_{hitung}	$p-value$
X_1	\hat{a}_0	-30730,663	-3,049	0,003
	\hat{a}_1	-4,181	-3,967	0,0001
	\hat{a}_2	557156,824	2,866	0,006
	\hat{a}_3	-14895,606	-2,887	0,005
X_2	\hat{a}_4	-2,615	-1,254	0,214
	\hat{a}_5	9993,462	1,668	0,100
	\hat{a}_6	-2753,908	-1,641	0,106
X_3	\hat{a}_7	0,883	0,823	0,413
	\hat{a}_8	45774,869	-3,047	0,003
	\hat{a}_9	11669,783	3,042	0,003

Berdasarkan hasil pada Tabel 4, terlihat bahwa ada 4 parameter yang tidak signifikan dan ada 6 parameter yang signifikan. Variabel X_1 memiliki parameter yang seluruhnya signifikan, sedangkan pada X_2 seluruh parameter tidak signifikan, dan pada X_3 terdapat 1 parameter yang tidak signifikan, sedangkan 2 parameter lainnya signifikan. Selanjutnya yaitu melakukan pengujian asumsi residual, yang pertama yaitu uji asumsi residual identik. Tabel 5 adalah hasil Uji *Glejser*.

Tabel 5. Hasil Uji *Glejser*

F_{hitung}	$p-value$
3,458	0,067

Berdasarkan Tabel 5, dapat diberikan keputusan bahwa F_{hitung} (3,458) $> F_{tabel(3,63;0,05)}$ (2,75) serta $p-value > \alpha$ yang berarti gagal tolak H_0 , sehingga dapat disimpulkan bahwa residual identik. Karena asumsi residual identik sudah terpenuhi, selanjutnya menguji asumsi independensi dengan *run test*. Hasil *run test* ada pada Tabel 6.

Tabel 6. Hasil *run test*

r	n_2	n_2	r_{atas}	r_{bawah}
36	33	34	42	11

Tabel 6 menunjukkan bahwa nilai r (36) $< r_{atas}$ (42) dan r (36) $> r_{bawah}$ (11) maka, gagal tolak H_0 dan dapat disimpulkan bahwa residualnya acak sehingga, tidak ada autokorelasi dan residual independen. Selanjutnya pengujian yang terakhir yaitu uji asumsi distribusi normal menggunakan uji *Kolmogorov Smirnov* yang hasilnya ada di Tabel 7.

Tabel 7. Hasil Uji *Kolmogorov Smirnov*

D	df	$p-value$
0,065	67	0,918

Tabel 7 menunjukkan nilai statistik uji sebesar 0,065 kurang dari $D_{(0,95;67)} = 0,166$ serta $p-value < \alpha$ yang artinya gagal tolak H_0 , sehingga residual berdistribusi normal.

5. KESIMPULAN

Berdasarkan hasil dari pemilihan osilasi optimal pada regresi nonparametrik deret *Fourier* dengan *Generalized Cross-Validation* (GCV) didapatkan model terbaik pada dua osilasi, yang mampu memberikan kinerja prediksi yang cukup baik. Model yang dihasilkan

menunjukkan bahwa secara simultan, ketiga variabel prediktor berpengaruh secara signifikan terhadap variabel respon. Walaupun secara parsial tidak semua parameter signifikan, tetapi model tetap dapat digunakan karena seluruh asumsi residual terpenuhi. Penerapan metode ini memberikan gambaran bahwa pendekatan deret *Fourier* mampu menangkap pola hubungan yang sifatnya berulang pada data kemiskinan yang dipengaruhi oleh beberapa variabel dari bidang ekonomi, pendidikan, dan ketenagakerjaan. Meskipun demikian, hasil analisis menunjukkan perlunya pengembangan lebih lanjut mengenai model dengan memberikan tambahan variabel prediktor atau mengombinasikan metode deret *Fourier* dengan pendekatan nonparametrik lain agar model yang dihasilkan dapat memberikan interpretasi yang lebih mendalam serta prediksi yang lebih akurat terhadap permasalahan kemiskinan.

DAFTAR PUSTAKA

- Amrullah M. N. & Hariksa Amalia S. (2022). Comparison of Generalized Cross Validation (GCV) Methods with Cross Validation (CV) to Determine Optimal Knots in Fourier Series Nonparametric Regression (Case Study: Poverty Rate in North Sumatra Province). *Jurnal Litbang Edusaintech*. 3(1):1-10. <http://journal.pwmjateng.com/index.php/jle> DOI: 10.51402/jle.v3i1.5
- BPS. (2011). *Penjelasan Data Kemiskinan*. Indonesia. Badan Pusat Statistik.
- BPS. (2024). *Statistik Indonesia 2024*. Indonesia. Badan Pusat Statistik.
- Daniel, W. W. (1989). *Statistika Nonparametrik Terapan*. Jakarta: Gramedia Pustaka Utama.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing 2nd Edition*. Texas. SRS Press.
- Fernandes, A. A. R. & Solimun. (2021). *Analisis Regresi dalam Pendekatan Fleksibel: Ilustrasi dengan Paket Program R*. Malang. Universitas Brawijaya Press.
- Laswinia, V. D. & Chamid, M. S. (2016). Analisis Pola Hubungan Persentase Penduduk Miskin dengan Faktor Lingkungan, Ekonomi, dan Sosial di Indonesia Menggunakan Regresi Spasial. *Jurnal Sains dan Seni ITS*. 5(2):235-240.
- Ni'matuzzahroh L, & Dani A. T. R. (2024). Nonparametric Regression Modeling with Multivariable Fourier Series Estimator on Average Length of Schooling in Central Java in 2023. *Inferensi*. 7(2):73-81. DOI: 10.12962/j27213862.v7i2.20219
- Octavanny, M. A. D., Budiantara, I. N., dan Ratnasari, V. (2017). Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Barat Menggunakan Regresi Semiparametrik Spline. *Jurnal Sains dan Seni ITS*. 6(1):122-128.
- Pynanjung P. A, Agustinus E, Junaidi J, Burhansyah R, & Oktoriana S. (2021). Poverty in the Indonesia-Malaysia border province (case study in West Kalimantan Province). *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah*. 9(5):401-412. DOI: 10.22437/ppd.v9i5.12760
- Saputro R. H. & Arif M. (2024). What factors affecting poverty rates in Indonesia? Empirical evidence from West Sumatera. *Journal of Enterprise and Development (JED)*. 6(1):248-258.
- Sari, E. P. & Novianti. (2024). Pengaruh PDRB terhadap Kemiskinan di Kalimantan Barat Tahun 2017-2022. *Ekodesinasi*. 2(1):36–56.
- Sari, R. S. & Budiantara, I. N. (2012). Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel. *Jurnal Sains dan Seni ITS*. 1(1):236-241.

- Suganda, A., Ramadhini, M., Harahap, P., Nugrahadi, E., & Gultom, G. (2024). The Effect of HDI and Unemployment Rate on Poverty in the Riau Islands. *Equity: Jurnal Ekonomi*. 8(1):58-65. DOI: 10.33019/equity.v%vi%i.308
- Utami T. W, Haris M. A, Prahutama A, & Purnomo E. A. (2020). Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods. In: *Journal of Physics: Conference Series*. Institute of Physics Publishing.
- Wahba G. (1990). *Spline Models for Observational Data*. Philadelphia. University of Wisconsin at Madison.
- Wu, H. & Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. New York. John Wiley & Sons.