

PEMODELAN *PENALIZED SPLINE* FAKTOR SOSIAL EKONOMI TERHADAP KEMISKINAN DI JAWA TIMUR

PENALIZED SPLINE MODELING OF SOCIOECONOMIC FACTORS AFFECTING POVERTY IN EAST JAVA PROVINCE

Andrea Tri Rian Dani^{1§}, Fachrian Bimantoro Putra^{2,3}

¹ Program Studi Statistika, Jurusan Matematika, Fakultas MIPA, Universitas Mulawarman
[Email: andreatririandani@fmipa.unmul.ac.id]

² Program Studi Statistika, Jurusan Matematika, Fakultas MIPA, Universitas Mulawarman
[Email: putrafachrian9@gmail.com]

³ BNI Staff Banking, Future Relationship Manager, BNI, Indonesia
[Email: Fachrian.Putra@bni.co.id]

[§]*Corresponding Author*

Received 10th Apr 2025; Accepted 10th Nov 2025; Published 1st Dec 2025;

Abstrak

Penelitian ini bertujuan untuk menganalisis pengaruh faktor sosial ekonomi terhadap tingkat kemiskinan di Provinsi Jawa Timur menggunakan pendekatan *penalized spline regression*. Model ini dipilih karena mampu menangkap hubungan non-linear antara variabel prediktor dan respon secara fleksibel. Tiga variabel utama yang dianalisis adalah Tingkat Pengangguran Terbuka, Rata-rata Lama Sekolah, dan Laju Pertumbuhan Ekonomi. Evaluasi model dilakukan dengan membandingkan nilai Generalized Cross Validation (GCV), koefisien determinasi (R^2), dan waktu komputasi pada berbagai kombinasi titik knot dan parameter penalti (λ). Hasil menunjukkan bahwa model terbaik diperoleh dengan tiga titik knot dan λ sebesar 0,001, menghasilkan GCV terendah (3,073) dan R^2 tertinggi (90,92%). Pengujian hipotesis secara simultan dan parsial mengindikasikan bahwa Rata-rata Lama Sekolah merupakan faktor yang paling signifikan memengaruhi kemiskinan, dengan arah hubungan negatif yang menunjukkan bahwa peningkatan tingkat pendidikan berperan penting dalam menurunkan tingkat kemiskinan. Meskipun kompleksitas komputasi meningkat seiring bertambahnya jumlah knot, hasil prediksi model tetap stabil dan akurat. Temuan ini menegaskan bahwa pendekatan *penalized spline* efektif untuk menganalisis isu sosial ekonomi yang bersifat dinamis dan kompleks, serta dapat menjadi dasar pengambilan kebijakan yang lebih berbasis data.

Kata Kunci: Kemiskinan, *Penalized Spline*, Regresi Nonparametrik, Sosial Ekonomi, Non-Linear

Abstract

*This study aims to analyze the influence of socioeconomic factors on poverty levels in East Java Province using the *penalized spline regression* approach. This model was chosen for its ability to flexibly capture non-linear relationships between predictor and response variables. The three main predictors analyzed are the Open Unemployment Rate, Average Years of Schooling, and Economic Growth Rate. Model*

evaluation was conducted by comparing the values of Generalized Cross Validation (GCV), the coefficient of determination (R^2), and computational time across various combinations of knot points and penalty parameters (λ). The results indicate that the optimal model is obtained with three knot points and a λ value of 0.001, producing the lowest GCV (3.073) and the highest R^2 (90.92%). Simultaneous and partial hypothesis testing reveals that the Average Years of Schooling is the most significant factor affecting poverty, showing a negative relationship that suggests higher education levels play an important role in reducing poverty rates. Although computational complexity increases with the number of knots, the model's predictive performance remains stable and accurate. These findings confirm that the penalized spline approach is effective for analyzing dynamic and complex socioeconomic issues and can serve as a robust foundation for data-driven policymaking.

Keywords: *Poverty, Penalized Spline, Nonparametric Regression, Socioeconomic, Non-Linear*

1. Pendahuluan

Kemiskinan merupakan permasalahan multidimensional yang tidak hanya berkaitan dengan aspek ekonomi, tetapi juga sosial, pendidikan, dan pembangunan wilayah [1], [2]. Berdasarkan data Badan Pusat Statistik (BPS), persentase penduduk miskin di Provinsi Jawa Timur pada Maret 2023 tercatat sebesar 10,35%, atau setara dengan sekitar 4,08 juta jiwa. Meskipun angka ini menunjukkan tren penurunan dibandingkan tahun-tahun sebelumnya, kesenjangan antarwilayah dan kelompok masyarakat masih cukup lebar. Realitas ini menandakan bahwa kemiskinan tetap menjadi isu strategis yang perlu ditangani dengan pendekatan berbasis data dan kebijakan yang tepat sasaran (*evidence-based policy*) [3]. Terlebih lagi, pengentasan kemiskinan merupakan salah satu prioritas utama dalam visi Presiden terpilih, Prabowo Subianto, yang menargetkan peningkatan kesejahteraan rakyat secara menyeluruh di seluruh Indonesia .

Untuk mendukung kebijakan yang efektif, diperlukan pemahaman yang lebih dalam mengenai faktor-faktor sosial ekonomi yang

berkontribusi terhadap kemiskinan. Beberapa indikator yang sering dikaitkan dengan tingkat kemiskinan di antaranya adalah Tingkat Pengangguran Terbuka, Rata-rata Lama Sekolah, dan Laju Pertumbuhan Ekonomi. Hubungan antara variabel-variabel tersebut dengan tingkat kemiskinan cenderung tidak bersifat linier, sehingga pendekatan model konvensional seperti regresi linier dapat kurang mampu menangkap pola yang sebenarnya [4], [5].

Salah satu pendekatan yang relevan untuk mengkaji hubungan antar variabel sosial ekonomi dan kemiskinan adalah regresi nonparametrik [6], [7]. Berbeda dengan regresi parametrik yang mengharuskan asumsi bentuk fungsi tertentu (misalnya linier), regresi nonparametrik bersifat lebih fleksibel dan mampu menyesuaikan diri terhadap pola data yang kompleks atau tidak diketahui sebelumnya [8], [9]. Hal ini sangat berguna ketika hubungan antara variabel tidak mengikuti pola yang sederhana, seperti yang kerap ditemukan dalam data sosial ekonomi.

Dari berbagai metode regresi nonparametrik yang ada, *penalized spline* menjadi salah satu

yang cukup banyak digunakan, terutama karena mampu menjaga keseimbangan antara fleksibilitas model dan kestabilan estimasi [10], [11]. Metode ini membagi data ke dalam segmen-segmen menggunakan titik-titik knot, lalu membangun fungsi potongan (*piecewise polynomial*) yang halus. Namun, semakin banyak knot yang digunakan, model bisa menjadi terlalu kompleks dan rentan terhadap *overfitting*.

Penalized spline menambahkan penalti terhadap kekasaran kurva, biasanya melalui penalti pada turunan kedua dari fungsi *spline* [12]. Dengan pendekatan ini, model tetap cukup fleksibel untuk menangkap pola penting dalam data, tetapi tidak terlalu rumit hingga kehilangan kemampuan generalisasi [13], [14]. Dalam penelitian ini, *Generalized Cross Validation* (GCV) digunakan untuk menentukan besar penalti serta lokasi dan jumlah knot yang optimal. GCV dipilih karena secara teori terbukti optimal secara asimtotik dan dalam banyak studi sebelumnya menunjukkan performa yang baik dalam memilih model yang tepat tanpa perlu membagi data secara eksplisit [15], [16].

Dengan mempertimbangkan kompleksitas permasalahan kemiskinan dan keterbatasan pendekatan analisis konvensional, penelitian ini menjadi penting untuk dilakukan. Pemanfaatan regresi nonparametrik *penalized spline* memungkinkan identifikasi hubungan nonlinier antara faktor-faktor sosial ekonomi dan tingkat kemiskinan secara lebih akurat. Hasil analisis yang lebih representatif ini diharapkan dapat memberikan dasar ilmiah yang kuat bagi pemerintah daerah dalam merumuskan strategi

pengentasan kemiskinan yang berbasis bukti (*evidence-based policy*). Selain itu, pendekatan ini juga berpotensi menjadi model analisis alternatif bagi penelitian-penelitian sosial ekonomi di daerah lain yang menghadapi karakteristik kemiskinan serupa, sehingga temuan penelitian tidak hanya bersifat lokal, tetapi juga memiliki nilai strategis dan aplikatif secara nasional.

2. Landasan Teori

2.1 Penalized Spline

Penalized spline merupakan salah satu estimator dalam regresi nonparametrik. Estimator ini memadukan fleksibilitas *spline* dengan pengendalian kekasaran estimasi kurva regresi melalui penambahan penalti (λ) [13].

Dalam analisis, estimator ini akan membagi rentang data menjadi beberapa segmen menggunakan titik-titik knot, lalu menyambungkan potongan (*piecewise polynomial*) secara halus. Namun, semakin banyak titik knot yang digunakan, semakin kompleks model yang dihasilkan, yang berisiko menyebabkan model yang kompleks dan *overfitting*. Hal ini akan diminimalisir dengan memberikan penalti terhadap kekasaran fungsi [10], [17].

Model umum *penalized spline* univariabel dituliskan pada Persamaan (1).

$$y_i = f(x_i) + \varepsilon_i \quad (1)$$

Dimana $i = 1, 2, \dots, n$.

Selanjutnya $f(x)$ dituliskan sebagai:

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_d x^d + \sum_{k=1}^K \beta_k (x - K_k)_+^d \quad (2)$$

Dengan

$\beta_0, \beta_1, \dots, \beta_d, \beta_k$: parameter yang diestimasi

$(x - K_k)_+^d$: basis fungsi *spline*

ε_i : *error* IIDN $(0, \sigma^2)$

y_i : variabel respon

Penalti diletakkan pada kekasaran fungsi, dengan:

$$Penalti = \lambda \int [f''(x)]^2 dx \quad (3)$$

Penalti diperlukan untuk mengendalikan tingkat kelengkungan (kekasaran) fungsi *spline*, agar model tidak terlalu mengikuti fluktuasi data (*overfitting*). Dengan menambahkan penalti terhadap turunan kedua fungsi, model diarahkan untuk menghasilkan fungsi yang lebih halus dan stabil, sehingga mampu menangkap pola utama tanpa kehilangan generalisasi.

Estimasi dilakukan dengan meminimalkan fungsi obyektif pada Persamaan (4).

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int [f''(x)]^2 dx \quad (4)$$

Selanjutnya fungsi regresi *penalized spline* dalam kasus multivariabel dapat ditulis pada Persamaan (5).

$$f(x) = \beta_0 + f_1(x_1) + \dots + f_p(x_p) \quad (5)$$

Pada penelitian ini, semua fungsi *spline* untuk masing-masing variabel prediktor dikenai penalti yang sama sehingga:

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{j=1}^p \int [f_j''(x_j)]^2 dx_j \quad (6)$$

Misalkan

$\mathbf{y} \in \mathbb{R}^n$: vektor respon

$\mathbf{X} \in \mathbb{R}^{n \times (p+1)}$: matriks prediktor linear

$\mathbf{Z} \in \mathbb{R}^{n \times K}$: matriks basis fungsi *spline*

$\boldsymbol{\beta} \in \mathbb{R}^{p+1}$: parameter linear

$\boldsymbol{\delta} \in \mathbb{R}^K$: parameter *spline*

$\mathbf{M} = [\mathbf{X} \ \mathbf{Z}] \in \mathbb{R}^{n \times (p+1+K)}$: matriks gabungan

$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} \in \mathbb{R}^{n \times (p+1+K)}$: vektor parameter gabungan.

Estimasi parameter dilakukan dengan meminimalkan Persamaan (7).

$$\mathcal{L}(\boldsymbol{\eta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta}\|^2 + \lambda \boldsymbol{\delta}^T \mathbf{S} \boldsymbol{\delta} \quad (7)$$

Jika $\mathbf{Z} \in \mathbb{R}^{n \times K}$ adalah matriks basis fungsi *spline* maka

$$\mathbf{S} = \mathbf{D}^T \mathbf{D} \quad (8)$$

Dimana $\mathbf{D}$ adalah matriks diferensiasi orde ke-2 dari koefisien *spline* $\boldsymbol{\delta}$.

$$\int [f''(x)]^2 dx \approx \boldsymbol{\delta}^T \mathbf{S} \boldsymbol{\delta} \quad (9)$$

Persamaan (7) dapat di modifikasi menjadi Persamaan (10).

$$\mathcal{L}(\boldsymbol{\eta}) = \|\mathbf{y} - \mathbf{M}\boldsymbol{\eta}\|^2 + \boldsymbol{\eta}^T \mathbf{P} \boldsymbol{\eta} \quad (10)$$

Penalti hanya diberikan pada komponen *spline*, sehingga

$$\mathbf{P} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda \mathbf{S} \end{bmatrix} \quad (11)$$

Untuk memperoleh estimasi parameter $\hat{\eta}$ dilakukan dengan meminimumkan Persamaan (10). Adapun turunan pertama dari $\mathcal{L}(\eta)$ terhadap η dan disamadengankan null yaitu:

$$\frac{\partial L}{\partial \eta} = -2\mathbf{M}^T(\mathbf{y} - \mathbf{M}\eta) + 2\mathbf{P}\eta = \mathbf{0}$$

$$= (\mathbf{M}^T\mathbf{M} + \mathbf{P})\eta = \mathbf{M}^T\mathbf{y}$$

Sehingga diperoleh solusi estimasi parameter untuk $\hat{\eta}$ pada Persamaan (12).

$$\hat{\eta} = (\mathbf{M}^T\mathbf{M} + \mathbf{P})^{-1}\mathbf{M}^T\mathbf{y} \quad (12)$$

Maka

$$\hat{f}(x_i) = \mathbf{M}\hat{\eta} \rightarrow \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\delta}$$

2.3 Generalized Cross-Validation

Dalam regresi *penalized spline*, setelah diperoleh estimasi parameter $\hat{\eta}$, nilai prediksi $\hat{\mathbf{y}}$ dapat dinyatakan pada Persamaan (13).

$$\hat{\mathbf{y}} = \mathbf{M}\hat{\eta} = \mathbf{M}((\mathbf{M}^T\mathbf{M} + \mathbf{P})^{-1}\mathbf{M}^T\mathbf{y}) \quad (13)$$

Persamaan (13) menunjukkan bahwa $\hat{\mathbf{y}}$ adalah hasil transformasi linear dari $\mathbf{y}$, sehingga dapat dituliskan dalam bentuk:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \quad (14)$$

Dimana

$$\mathbf{H} = \mathbf{M}(\mathbf{M}^T\mathbf{M} + \mathbf{P})^{-1}\mathbf{M}^T \quad (15)$$

$\mathbf{H}$ adalah matriks hat yang digunakan untuk menunjukkan seberapa baik model *penalized spline*. Evaluasi yang biasa digunakan adalah *Generalized Cross-Validation* (GCV) pada Persamaan (16) [18], [19].

$$GCV(\lambda, \mathbf{K}) = \frac{1}{n} \frac{\|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2}{\left(\frac{1}{n} \text{trace}(\mathbf{I} - \mathbf{H})\right)^2} \quad (16)$$

2.4 Koefisien Determinasi

Koefisien determinasi (R^2) merupakan suatu metrik untuk mengukur seberapa besar proporsi variasi dalam variabel respon yang bisa dijelaskan dalam model [20], [21]. Dalam regresi *penalized spline*, formula R^2 dituliskan pada Persamaan (17).

$$R^2 = 1 - \frac{SSE}{SST} \rightarrow 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

Pemilihan koefisien determinasi (R^2) sebagai ukuran kebaikan model dilakukan karena metrik ini mampu memberikan gambaran yang jelas mengenai seberapa besar variasi Persentase Penduduk Miskin yang dapat dijelaskan oleh variabel prediktor dalam model *penalized spline*.

2.5 Data dan Sumber Data

Data pada penelitian ini merupakan data sekunder dari publikasi Badan Pusat Statistik Tahun 2023. Unit observasi adalah Kabupaten/Kota di Provinsi Jawa Timur. Teknik sampling yang digunakan adalah *purposive sampling* dengan mempertimbangkan keterbaruan serta ketersediaan data.

Variabel yang digunakan merupakan variabel respon yaitu Persentase Penduduk Miskin dengan tiga variabel prediktor yang diduga memengaruhi,

yaitu Tingkat Pengangguran Terbuka, Rata-rata Lama Sekolah, dan Laju Pertumbuhan Ekonomi.

2.6 Tahapan Penelitian

Tahapan penelitian dituliskan dalam 3 garis besar yaitu:

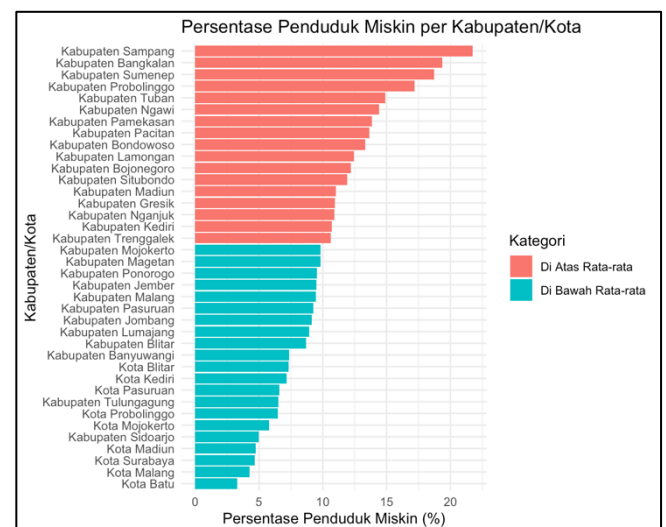
1. Eksplorasi Data menggunakan *Bar Plot* dan *Scatter Plot* untuk mengidentifikasi pola hubungan antara variabel prediktor dengan respon.
2. Pemodelan dengan *Penalized Spline*
 - a. Pemilihan titik knot sebanyak 1,2, dan 3 pada setiap variabel prediktor. Titik knot dipilih secara sistematis dari sebaran data setiap prediktor menggunakan kombinasi.
 - b. Penambahan *penalty* diletakkan pada bagian spline. Pada penelitian ini, semua fungsi *spline* untuk masing-masing variabel prediktor dikenai *penalty* yang sama
 - c. Estimasi parameter model regresi *penalized spline* sesuai pada Persamaan (12)
 - d. Nilai λ yang dicobakan pada rentang 0,001 sampai 100 dalam skala logaritmik, yaitu:
$$\lambda \in \{10^{-3}, \dots, 10^2\}$$
 - e. Model regresi *penalized spline* terbaik dipilih berdasarkan nilai GCV minimum.
 - f. Pengujian hipotesis pada model regresi *penalized spline* terbaik secara simultan dan parsial.
 - g. Visualisasi hasil

3. Analisis *running time* dimaksudkan untuk menilai efisiensi model dengan melihat waktu eksekusi model dengan jumlah knot yang berbeda.

3. Hasil Dan Pembahasan

3.1 Eskplorasi Data

Eskplorasi data adalah bagian penting dalam proses pemodelan untuk melihat gambaran umum dari data yang akan dianalisis. Pada penelitian ini eksplorasi yang digunakan menggunakan *Bar Plot* dan *Scatter Plot*.

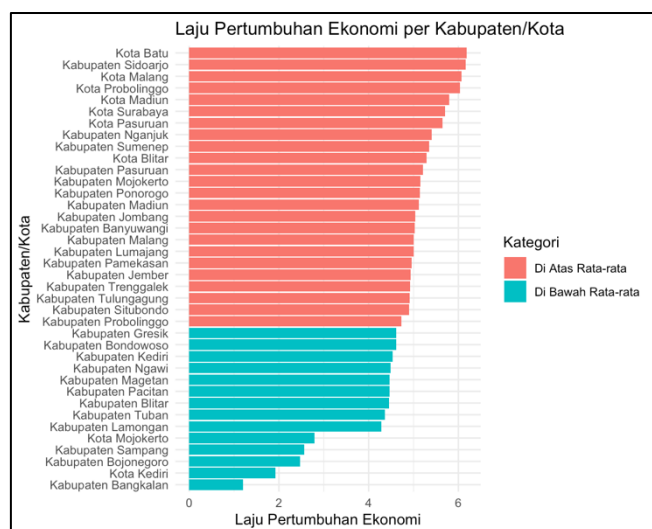


Gambar 1. *Bar Plot* untuk Persentase Penduduk Miskin

Gambar 1 memperlihatkan perbedaan tingkat kemiskinan di berbagai kabupaten dan kota di Jawa Timur. Terlihat jelas bahwa beberapa wilayah seperti Kabupaten Sampang, Bangkalan, dan Sumenep memiliki persentase penduduk miskin yang jauh lebih tinggi dibandingkan daerah lainnya. Sampang bahkan mendekati angka 21%, menjadikannya wilayah dengan tingkat kemiskinan tertinggi. Sebaliknya, kota-kota besar seperti Kota Batu, Kota Malang, dan

Surabaya menunjukkan angka kemiskinan yang jauh lebih rendah dari rata-rata.

Perbedaan ini menunjukkan bahwa belum semua wilayah merasakan hasil pembangunan secara merata. Daerah-daerah dengan kemiskinan tinggi umumnya memiliki tantangan lebih besar dalam hal akses pendidikan, kesehatan, dan peluang ekonomi. Sementara itu, kota-kota besar yang lebih maju menikmati fasilitas dan akses yang lebih baik. Kesenjangan ini menjadi pengingat bahwa upaya pengentasan kemiskinan tidak bisa disamaratakan. Perlu kebijakan yang peka terhadap kondisi lokal dan berbasis data yang akurat. Di

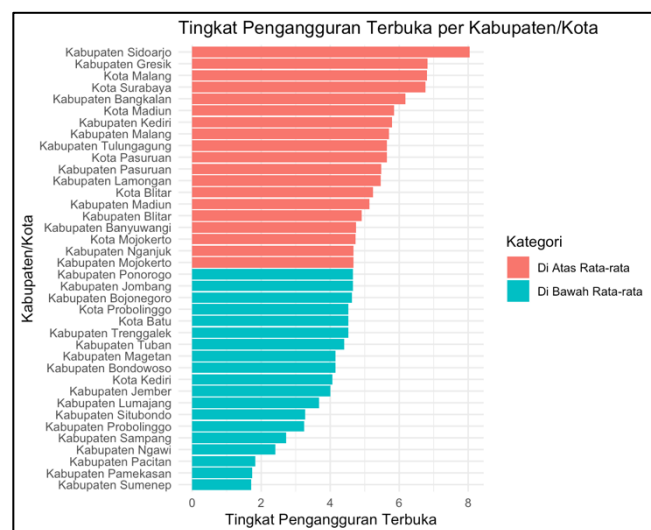


Gambar 2. Bar Plot untuk Laju Pertumbuhan Ekonomi

Gambar 2 menampilkan laju pertumbuhan ekonomi per kabupaten/kota di Jawa Timur. Kota Batu menempati posisi teratas dengan pertumbuhan ekonomi tertinggi, diikuti oleh Kota Sidoarjo, Kota Malang, dan kota-kota besar lainnya seperti Surabaya dan Probolinggo. Sebagian besar wilayah dengan pertumbuhan tinggi merupakan daerah perkotaan atau kawasan

penyangga kota besar yang memiliki akses infrastruktur, investasi, dan pasar yang lebih baik.

Di sisi lain, kabupaten-kabupaten seperti Bangkalan, Sampang, dan Pamekasan menunjukkan laju pertumbuhan ekonomi yang rendah, bahkan di bawah 2%. Hal ini mencerminkan adanya kesenjangan pertumbuhan antar wilayah di Jawa Timur. Wilayah dengan pertumbuhan ekonomi yang lambat umumnya adalah daerah pedesaan atau kepulauan yang menghadapi tantangan dalam konektivitas, investasi, dan pengembangan sektor ekonomi produktif.



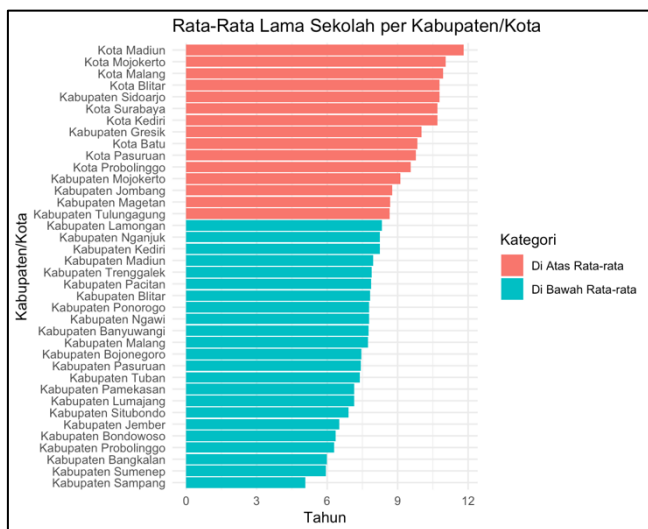
Gambar 3. Bar Plot untuk Tingkat Pengangguran Terbuka

Gambar 3 memperlihatkan tingkat pengangguran terbuka di berbagai kabupaten dan kota di Jawa Timur. Terlihat bahwa Kabupaten Sidoarjo menempati posisi teratas dengan tingkat pengangguran tertinggi, disusul oleh Kabupaten Gresik, Kota Malang, dan Kota Surabaya. Umumnya, daerah-daerah dengan pengangguran tinggi ini merupakan kawasan industri dan urban, di mana dinamika pasar kerja sangat cepat,

namun tidak selalu diimbangi dengan ketersediaan lapangan kerja yang cukup.

Sebaliknya, wilayah seperti Kabupaten Sumenep, Pamekasan, dan Pacitan menunjukkan tingkat pengangguran yang relatif rendah. Hal ini bisa jadi karena masyarakat di daerah tersebut lebih banyak bekerja di sektor informal atau pertanian, yang meskipun tidak selalu menghasilkan pendapatan tinggi, namun relatif menyerap lebih banyak tenaga kerja.

Kesenjangan tingkat pengangguran ini mencerminkan tantangan yang dihadapi setiap daerah dalam menyesuaikan antara ketersediaan dan kebutuhan tenaga kerja. Kota-kota besar menghadapi tantangan mismatch keterampilan, sementara daerah pedesaan berhadapan dengan keterbatasan kesempatan kerja formal.



Gambar 4. Bar Plot untuk Rata-Rata Lama Sekolah

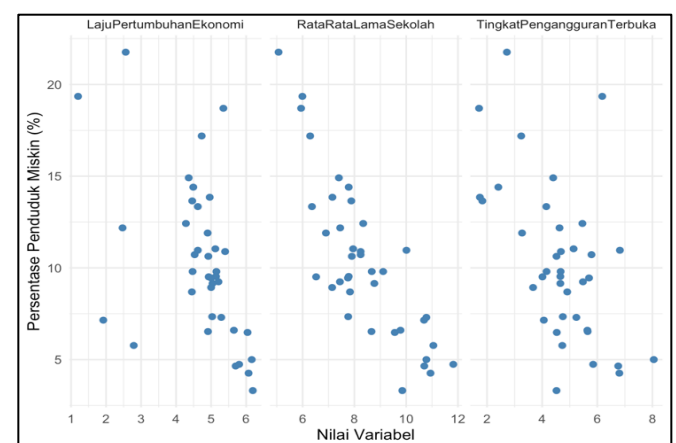
Gambar 4 menunjukkan rata-rata lama sekolah penduduk di berbagai kabupaten/kota di Jawa Timur. Dari visualisasi ini terlihat bahwa Kota Madiun, Kota Mojokerto, dan Kota Malang menempati posisi tertinggi, dengan rata-rata lama sekolah mendekati atau lebih dari 10 tahun. Ini mencerminkan akses pendidikan yang relatif

lebih baik dan mungkin juga mencerminkan tingginya kesadaran masyarakat terhadap pentingnya pendidikan di wilayah-wilayah tersebut.

Sebaliknya, beberapa daerah seperti Kabupaten Sampang, Bangkalan, dan Pamekasan masih memiliki rata-rata lama sekolah di bawah 7 tahun. Kondisi ini bisa disebabkan oleh berbagai faktor, seperti keterbatasan fasilitas pendidikan, tantangan ekonomi keluarga yang membuat anak harus bekerja lebih awal, atau rendahnya tingkat partisipasi pendidikan menengah dan atas.

Kesenjangan ini menjadi cerminan nyata bahwa belum semua masyarakat di Jawa Timur memperoleh kesempatan belajar yang sama. Padahal, pendidikan yang baik merupakan salah satu fondasi penting untuk keluar dari kemiskinan.

Selanjutnya pola hubungan antara variabel prediktor terhadap variabel respon diidentifikasi menggunakan scatter plot yang ditampilkan pada Gambar 5.



Gambar 5. Hubungan variabel prediktor terhadap variabel respon

Scatter plot pada Gambar 5 memperlihatkan bagaimana beberapa faktor sosial ekonomi yakni

laju pertumbuhan ekonomi, rata-rata lama sekolah, dan tingkat pengangguran terbuka berkaitan dengan persentase penduduk miskin di berbagai kabupaten dan kota di Jawa Timur. Untuk variabel laju pertumbuhan ekonomi, terlihat bahwa daerah dengan pertumbuhan tinggi tidak selalu memiliki tingkat kemiskinan yang rendah. Ini menunjukkan bahwa pertumbuhan ekonomi saja belum tentu menjamin pemerataan kesejahteraan. Bisa jadi, pertumbuhan tersebut hanya dinikmati oleh kelompok tertentu tanpa menyentuh masyarakat rentan.

Pada variabel rata-rata lama sekolah, terlihat kecenderungan yang lebih jelas: semakin tinggi rata-rata lama sekolah, umumnya tingkat kemiskinan lebih rendah. Artinya, pendidikan tampaknya masih menjadi salah satu benteng utama dalam mengurangi risiko kemiskinan.

Sementara itu, hubungan antara tingkat pengangguran terbuka dengan kemiskinan tampak cukup bervariasi. Beberapa daerah dengan tingkat pengangguran tinggi memang mengalami kemiskinan yang tinggi, tapi ada pula yang sebaliknya. Pola-pola yang muncul ini menggambarkan bahwa hubungan antara variabel sosial ekonomi dan kemiskinan tidak bisa disederhanakan begitu saja. Karena itulah pendekatan statistik yang lebih fleksibel, seperti regresi nonparametrik *penalized spline*, menjadi penting untuk digunakan. Metode ini memungkinkan kita melihat pola yang lebih halus dan realistis, yang bisa menjadi dasar dalam menyusun kebijakan pengentasan kemiskinan yang lebih tepat dan efektif.

3.2 Pemodelan Kemiskinan dengan *Penalized Spline*

Pemodelan dengan *penalized spline* pada penelitian ini menggunakan satu hingga tiga titik knot. Jumlah titik knot yang digunakan sama untuk setiap variabel prediktor. Hasil pemodelan ditampilkan pada Tabel 1 hingga Tabel 3.

Tabel 1. Pemodelan dengan Satu Titik Knot

λ	Titik Knot			GCV	R ²
	X ₁	X ₂	X ₃		
1,438	3,677	7,165	2,748	3,866	81,821
0,785	3,677	7,165	2,748	3,879	83,086
0,428	3,677	7,165	2,748	3,914	83,970

Tabel 1 menjelaskan jika model dengan satu titik knot menghasilkan nilai GCV antara 3,866–3,914 dan R² sebesar 81,82%–83,97%.

Tabel 2. Pemodelan dengan Dua Titik Knot

λ	Titik Knot			GCV	R ²
	X ₁	X ₂	X ₃		
0,001	4,114	7,630	3,093	3,242	89,297
	4,333	7,863	3,265		
0,001	4,114	7,630	3,093	3,252	88,729
	4,333	7,863	3,265		
0,004	4,114	7,630	3,093	3,294	88,070
	4,333	7,863	3,265		

Tabel 2 menjelaskan jika penambahan satu titik knot meningkatkan performa model dengan GCV 3,242–3,294 dan R² mencapai 88,07%–89,30%. Nilai λ kecil (0,001–0,004) memberikan keseimbangan antara kelengkungan dan kestabilan model.

Tabel 3. Pemodelan dengan Tiga Titik Knot

λ	Titik Knot			GCV	R ²
	X ₁	X ₂	X ₃		
0,001	3,677	7,164	2,748	3,073	90,508
	3,896	7,397	2,920		
	4,115	7,630	3,093		
0,001	3,458	6,932	2,576	3,089	90,916
	3,896	7,397	2,920		
	4,114	7,630	3,093		

	3,677	7,164	2,748		
0,002	3,896	7,397	2,920	3,092	90,845
	4,333	7,863	3,265		

Berdasarkan Tabel 3, model dengan tiga titik knot memberikan hasil terbaik, ditunjukkan oleh GCV terendah ($\pm 3,08$) dan R^2 tertinggi ($\pm 90,9\%$). Model ini paling optimal dalam menangkap hubungan nonlinier antarvariabel tanpa overfitting.

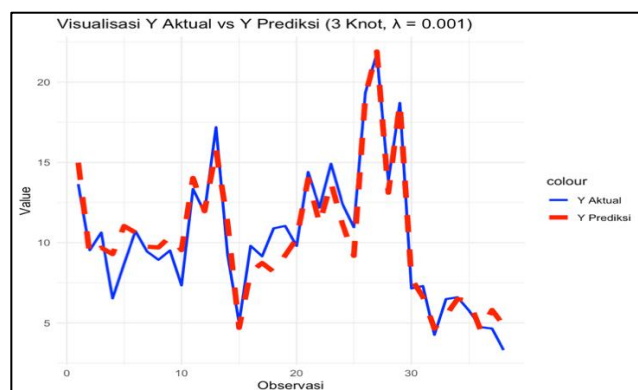
Dari hasil pemodelan *penalized spline* dengan variasi jumlah titik knot, kita bisa melihat bagaimana kompleksitas model memengaruhi akurasi dan kesesuaian model terhadap data. Terdapat tiga skenario yang diuji: menggunakan satu, dua, dan tiga titik knot, masing-masing dengan kombinasi parameter penalti (λ) dan lokasi knot yang berbeda-beda.

Pada pemodelan dengan satu titik knot, hasilnya cukup baik, dengan nilai GCV sekitar 3,866 dan R^2 mencapai 83,97%. Ini menunjukkan bahwa model sudah bisa menjelaskan sebagian besar variasi data, tetapi masih ada ruang untuk perbaikan.

Ketika jumlah titik knot ditambah menjadi dua, performa model meningkat. GCV turun ke angka 3,242 dan R^2 naik menjadi 89,30%. Ini artinya, model semakin mampu menangkap pola hubungan antara faktor-faktor sosial ekonomi dan tingkat kemiskinan.

Namun, hasil terbaik diperoleh saat digunakan tiga titik knot. Model ini menghasilkan GCV terendah sebesar 3,073 dan nilai R^2 tertinggi, yaitu 90,92%. Secara keseluruhan, bisa disimpulkan bahwa model dengan tiga titik knot dan nilai λ sebesar 0,001 adalah model paling optimal. Model ini seimbang serta cukup fleksibel

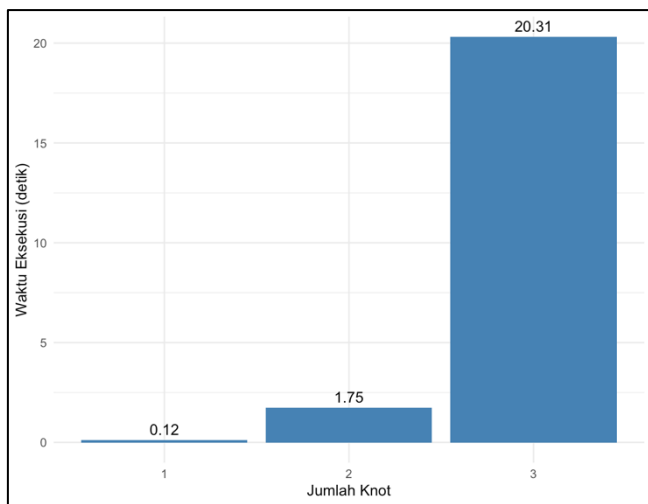
untuk menangkap pola data yang kompleks, tapi juga cukup stabil untuk menghasilkan estimasi yang dapat diandalkan.



Gambar 6. Visualisasi y dan $\hat{y}$ model terbaik

Gambar 6 menunjukkan perbandingan antara nilai aktual (garis biru) dan nilai prediksi (garis merah) dari persentase penduduk miskin berdasarkan model *penalized spline* terbaik, yaitu model dengan tiga titik knot dan nilai lambda (λ) sebesar 0,001.

Secara visual, garis prediksi mengikuti pola garis aktual dengan cukup baik. Fluktuasi atau perubahan naik-turun pada data asli berhasil ditangkap oleh model tanpa terlihat terlalu kaku. Hal ini menunjukkan bahwa model memiliki keseimbangan yang baik antara fleksibilitas dan kestabilan. Meskipun terdapat beberapa perbedaan antara prediksi dan data aktual, khususnya pada beberapa titik ekstrem, secara umum model mampu merepresentasikan tren dan pola utama dari data kemiskinan di setiap observasi.

Gambar 7. *Running Time Model Penalized Spline*

Gambar 6 menunjukkan waktu eksekusi (*running time*) dari model *spline* berdasarkan jumlah titik knot yang digunakan. Terlihat bahwa semakin banyak jumlah knot, waktu yang dibutuhkan untuk menjalankan model meningkat secara signifikan. Model dengan satu knot hanya memerlukan sekitar 0,12 detik, dua knot memerlukan sekitar 1,75 detik, sedangkan model dengan tiga knot memakan waktu lebih dari 20 detik. Hal ini menggambarkan bahwa semakin kompleks struktur model dalam hal ini ditandai oleh jumlah knot semakin tinggi pula beban komputasinya. Meskipun model dengan tiga knot memberikan hasil prediksi yang paling akurat, seperti ditunjukkan oleh nilai R^2 dan GCV yang optimal, namun konsekuensinya adalah peningkatan kompleksitas komputasi yang cukup tajam.

3.3 Pengujian Hipotesis Model Penalized Spline Terbaik

Pengujian hipotesis dilakukan secara simultan dan parsial. Tabel ANOVA ditampilkan pada Tabel 4.

Tabel 4. ANOVA untuk *Penalized Spline* Terbaik

Sumber	df	SS	MS	F-Hit	P Value
Regresi	12	625,3442	52,112	35,1814	0,0000
Error	25	56,2871	1,482		
Total	37	690,9209			

Hasil uji ANOVA menunjukkan bahwa model *penalized spline* yang dibangun dengan tiga titik knot mampu secara signifikan menjelaskan variasi dalam data. Dengan nilai F-hitung sebesar 35,18 dan p-value yang sangat kecil ($p < 0,05$), model ini terbukti secara statistik signifikan. Artinya, kombinasi dari ketiga variabel prediktor: laju pertumbuhan ekonomi, rata-rata lama sekolah, dan tingkat pengangguran terbuka memiliki hubungan yang signifikan terhadap persentase penduduk miskin di wilayah yang dikaji.

Tabel 5. Uji Parsial

Variabel	Parameter	Estimasi	T-Hit	P Value
	β_0	62,6720	10,1516	0,0000
Tingkat Pengangguran Terbuka	β_1	-1,7425	-2,7043	0,0121*
	β_{11}	-21,1163	-1,7710	0,0887
	β_{12}	41,6190	1,7309	0,0958
	β_{13}	-18,6275	-1,4425	0,1616
Rata-rata Lama Sekolah	β_2	-5,4169	-6,5316	0,0000*
	β_{21}	19,8544	2,8307	0,0090*
	β_{22}	-24,7018	-2,4639	0,0210*
	β_{23}	9,5044	2,1306	0,0431*
Laju Pertumbuhan Ekonomi	β_3	-3,3457	-2,4622	0,0211*
	β_{31}	17,3227	0,3723	0,7128
	β_{32}	8,8246	0,1002	0,9210
	β_{33}	-25,4391	-0,5913	0,5596

Berdasarkan hasil estimasi parameter dari model *penalized spline*, dapat dilihat bahwa tidak semua koefisien untuk masing-masing variabel prediktor signifikan secara statistik. Namun, ini justru menunjukkan salah satu kelebihan utama dari pendekatan *spline* dibandingkan model

parametrik biasa. Dalam *spline*, sebuah variabel tetap dapat dianggap berpengaruh secara signifikan terhadap respon apabila setidaknya salah satu dari komponen (koefisien) spline-nya menunjukkan signifikansi statistik. Model regresi penalized spline dengan titik-titik knot berdasarkan Tabel 3 yang optimal sebagai berikut:

$$\begin{aligned}\hat{y} = & 62,6720 - 1,7425X_1 - 21,1163(X_1 - K_1)_+ \\ & + 41,6190(X_1 - K_2)_+ - 18,6275(X_1 - K_3)_+ \\ & - 5,4169X_2 + 19,8544(X_2 - K_1)_+ \\ & - 24,7018(X_2 - K_2)_+ + 9,5044(X_2 - K_3)_+ \\ & - 3,3457X_3 + 17,3227(X_3 - K_1)_+ \\ & + 8,8246(X_3 - K_2)_+ - 25,4391(X_3 - K_3)_+\end{aligned}$$

Interpretasi:

1. Intersep ($\beta_0 = 62,6720$)

Nilai dasar dari variabel respon (y) Ketika seluruh variabel prediktor berada pada nilai minimum (nol). Nilai ini signifikan secara statistik (P Value = 0,0000).

2. Tingkat Pengangguran Terbuka (X_1)

- Koefisien $\beta_1 = -1,7425$ menunjukkan pengaruh negatif signifikan terhadap variabel respon (y), artinya peningkatan 1% pengangguran terbuka dapat menurunkan kemiskinan sekitar 1,74 satuan.
- Koefisien spline β_{11} , β_{12} , dan β_{13} menunjukkan hubungan nonlinear antara pengangguran dan kemiskinan. Pola menunjukkan fluktuasi lokal terhadap perubahan Tingkat pengangguran pada titik-titik tertentu (knot 1 hingga ke-3).

3. Rata-rata Lama Sekolah (X_2)

- Koefisien $\beta_2 = -5,4169$ menunjukkan bahwa peningkatan rata-rata lama sekolah berpengaruh menurunkan tingkat kemiskinan.
- Koefisien spline β_{21} , β_{22} , dan β_{23} menunjukkan hubungan nonlinear antara pendidikan dan kemiskinan. Pola menunjukkan fluktuasi lokal terhadap perubahan rata-rata lama sekolah pada titik-titik tertentu (knot 1 hingga ke-3).

4. Laju Pertumbuhan Ekonomi (X_3)

- Koefisien $\beta_3 = -3,3457$ menunjukkan bahwa peningkatan pertumbuhan ekonomi berpengaruh menurunkan tingkat kemiskinan.
- Koefisien spline β_{31} , β_{32} , dan β_{33} menunjukkan hubungan nonlinear antara pertumbuhan ekonomi dan kemiskinan. Pola menunjukkan fluktuasi lokal terhadap perubahan pertumbuhan ekonomi pada titik-titik tertentu (knot 1 hingga ke-3).

Dengan pendekatan *penalized spline*, diperoleh gambaran yang lebih kaya dan fleksibel mengenai hubungan antara variabel-variabel prediktor dengan kemiskinan. Tidak hanya melihat pada signifikansi satu parameter tunggal, tapi juga pada bentuk kontribusi variabel secara keseluruhan, yang sangat berguna dalam konteks data sosial ekonomi yang dinamis dan tidak selalu mengikuti pola linear sederhana.

4. Kesimpulan Dan Saran

Analisis menunjukkan bahwa kemiskinan di Jawa Timur masih menjadi isu serius yang dipengaruhi secara signifikan oleh faktor sosial ekonomi seperti tingkat pengangguran terbuka, rata-rata lama sekolah, dan laju pertumbuhan ekonomi. Dengan menggunakan model *penalized spline*, hubungan non-linear antar variabel dapat ditangkap secara lebih fleksibel dan akurat. Model terbaik diperoleh dengan tiga titik knot dan penalti $\lambda = 0,001$, menghasilkan nilai GCV terendah dan R^2 tertinggi sebesar 90,916. Rata-rata lama sekolah terbukti menjadi faktor paling berpengaruh, sehingga peningkatan akses dan kualitas pendidikan menjadi rekomendasi utama. Model *penalized spline* juga menunjukkan bahwa kompleksitas komputasi meningkat seiring jumlah knot, sehingga efisiensi perlu dipertimbangkan. Pendekatan *penalized spline* layak dijadikan alat bantu dalam perumusan kebijakan berbasis data di bidang sosial ekonomi.

Daftar Pustaka

- [1] Y. Wang, S. Jia, W. Qi, and C. Huang, "Examining Poverty Reduction of Poverty-Stricken Farmer Households under Different Development Goals: A Multiobjective Spatio-Temporal Evolution Analysis Method," *Int J Environ Res Public Health*, vol. 19, no. 19, Oct. 2022, doi: 10.3390/ijerph191912686.
- [2] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, "Modeling Poverty Percentages in the Papua Islands using Fourier Series in Nonparametric Regression Multivariable," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1397/1/012071.
- [3] V. Ratnasari, S. H. Audha, and A. T. R. Dani, "Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression," Dec. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.mex.2023.102379.
- [4] N. Budiantara, M. Ratna, I. Zain, and W. Wibowo, "Modeling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach," *International Journal of Basic & Applied Sciences IJBAS-IJENS*, vol. 12, no. 06, pp. 119–124, 2012.
- [5] I. N. Budiantara *et al.*, "Modeling Percentage of Poor People In Indonesia Using Kernel And Fourier Series Mixed Estimator In Nonparametric Regression," *REVISTA INVESTIGACION OPERACIONAL*, vol. 40, no. 4, pp. 538–550, 2019.
- [6] Y. Andriyana, J. Suprijadi, Y. Suparman, S. Winarni, and I. G. N. M. Jaya, "Modelling the percentage of poverty based on an open unemployment rate using some nonparametric regression techniques," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1742-6596/1265/1/012023.
- [7] V. Ratnasari, I. N. Budiantara, I. Zain, M. Ratna, and N. P. A. M. Mariati, "Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models (Application- Data of Poverty in Papua, Indonesia)," *International Journal of Basic & Applied Sciences (IJBAS-IJENS)*, vol. 15, no. 04, pp. 9–12, 2015.
- [8] A. T. R. Dani, L. Ni'matuzzahroh, V. Ratnasari, and I. N. Budiantara, "Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal," *Inferensi*, vol. 4, no. 1, p. 47, Mar. 2021, doi: 10.12962/j27213862.v4i1.8737.
- [9] A. T. R. Dani and L. Ni'matuzzahroh, "Penerapan Keluarga Model Spline Truncated Polinomial pada Regresi Nonparametrik," *Inferensi*, vol. 5, no. 1, p. 37, 2022, doi: 10.12962/j27213862.v5i1.12537.

- [10] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "An extended model of penalized spline with the addition of Kernel Functions in nonparametric regression model," *Applied Mathematics and Information Sciences*, vol. 13, no. 3, pp. 453–460, 2019, doi: 10.18576/amis/130318.
- [11] M. Roozbeh, M. Maanavi, and A. Mohamed, "Penalized least squares optimization problem for high-dimensional data," *Int. J. Nonlinear Anal. Appl.*, vol. 14, pp. 2008–6822, 2023, doi: 10.22075/ijnaa.2022.26998.3470.
- [12] A. A. R. Fernandes, I. N. Budiantara, and B. W. Otok, "Reproducing Kernel Hilbert space and penalized weighted least square in nonparametric regression," *Applied Mathematical Sciences*, vol. 8, no. 145–148, pp. 7289–7300, 2014, doi: 10.12988/ams.2014.49759.
- [13] A. Islamiyati and N. Chamidah, "Penalized spline estimator with multi smoothing parameters in bi-response multi-predictor nonparametric regression model for longitudinal data," *Songklanakar J. Sci. Technol.*, vol. 42, no. 4, pp. 897–909, 2020.
- [14] H. Becher, G. Kauerniann, P. Khomski, and B. Kouyaté, "Using penalized splines to model age and season-of-birth-dependent effects of childhood mortality risk factors in rural burkina faso," *Biometrical Journal*, vol. 51, no. 1, pp. 110–122, 2009, doi: 10.1002/bimj.200810496.
- [15] V. Ratnasari, I. N. Budiantara, and A. T. R. Dani, "Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) Methods," *Int J Adv Sci Eng Inf Technol*, vol. 11, no. 6, pp. 2400–2406, 2021.
- [16] A. T. R. Dani, V. Ratnasari, and I. N. Budiantara, "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression," *IOP Conf Ser Mater Sci Eng*, vol. 1115, no. 1, p. 012020, Mar. 2021, doi: 10.1088/1757-899x/1115/1/012020.
- [17] L. N. Berry and N. E. Helwig, "Cross-Validation, Information Theory, or Maximum Likelihood? A Comparison of Tuning Methods for Penalized Splines," *Stats (Basel)*, vol. 4, no. 3, pp. 701–724, Sep. 2021, doi: 10.3390/stats4030042.
- [18] A. T. R. Dani and N. Y. Adrianingsih, "Pemodelan Regresi Nonparametrik dengan Estimator Spline Truncated vs Deret Fourier," *Jambura Journal of Mathematics*, vol. 3, no. 1, pp. 26–36, 2021, doi: 10.34312/jjom.v3i1.7713.
- [19] N. Y. Adrianingsih and A. T. R. Dani, "Estimasi Model Regresi Semiparametrik Spline Truncated Menggunakan Metode Maximum Likelihood Estimation (MLE)," *Jambura Journal of Probability and Statistics*, vol. 2, no. 2, pp. 56–63, 2021, doi: 10.34312/jjps.v2i2.10255.
- [20] A. T. R. Dani, N. Y. Adrianingsih, and A. Ainurrochmah, "Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi (Studi Kasus: Gini Ratio Kabupaten/Kota Di Provinsi Jawa Timur Tahun 2017)," *Jambura Journal of Probability and Statistics*, vol. 1, no. 2, pp. 99–106, 2020.
- [21] V. Ratnasari, S. H. Utama, and A. T. R. Dani, "Toward Sustainable Development Goals (SDGs) with Statistical Modeling: Recursive Bivariate Binary Probit," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 8, pp. 1515–1521, 2024.