

Modeling of Employment Participation Rate Against the Percentage of Poor Population in East Java in 2023 Using B-Spline Method

Gilang Ibnul Farizi, Zilrahmi*, Dony Permana, Admi Salma

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: zilrahmi@fmipa.unp.ac.id

Submitted : 02 Agustus 2024

Revised : 12 Agustus 2024

Accepted : 13 Agustus 2024

ABSTRACT

Poverty is a common issue in Indonesia.. The greatest percentage of poverty in East Java Province in 2023 was 21,760, according to data on the percentage of the poor population versus the labor force participation rate (LFPR) each district/city, which consisted of 38 districts/cities in the province. The best way to reduce poverty is deemed to be through employment.. The data in this study shows a distribution pattern that does not form a specific pattern, making it difficult to analyze using parametric methods. Therefore, the appropriate approach is Nonparametric Regression. In this study, the nonparametric regression used is the B-Spline regression model. The suitability of the model is based on the Mean Squared Error (MSE) value of the model. The analysis results indicate that the B-Spline regression model achieves an MSE value of 20.11447. The optimal MSE value is obtained from B-Spline estimation with order 2. This suggests that the B-Spline method provides a good explanation in addressing the issue.

Keywords: B-Spline, Labor Force Participation Rate (LFRP), Percentage of Poor People.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kemiskinan didefinisikan sebagai keadaan ketika masyarakat tidak dapat melakukan pemenuhan kebutuhan sehari-hari yang layak (Bank, 2004). Faktor yang menjadi penyebab kemiskinan umumnya disebabkan oleh kurangnya pendapatan dan aset. Akibatnya kebutuhan dasar manusia seperti makanan, pakaian, perumahan, layanan kesehatan, serta Pendidikan yang layak. Hal ini bermuara dengan meningkatnya pengangguran. Tingginya pengangguran akan mengurangi pendapatan yang diterima masyarakat beserta tingkat kemakmuran yang telah dicapai. Kemiskinan, masalah tambahan, akan muncul sebagai akibat dari penurunan tingkat kemakmuran (Sukirno, 2000).

Masalah pengangguran di suatu wilayah menjadi semakin serius ketika cepatnya pertumbuhan Angkatan kerja tidak diiringi dengan perkembangan lapangan kerja. Tingginya tingkat pengangguran menandakan kurangnya pembangunan di suatu negara. Kemiskinan dapat dipengaruhi oleh pengangguran dalam berbagai cara (Tambunan, 2001).

Suatu model regresi dapat digunakan untuk menunjukkan hubungan antara tingkat partisipasi angkatan kerja dengan persentase kemiskinan. Regresi parametrik adalah teknik pendugaan model regresi yang bergantung pada persamaan regresi dengan jenis tertentu, seperti linier, kuadrat, dll. Akan tetapi, karena seringkali ada masalah untuk menemukan model melalui sebaran data, regresi nonparametrik digunakan. Regresi spline adalah salah satu metode estimasi regresi nonparametrik yang memperhitungkan penghalusan. Regresi spline memiliki kelebihan dalam keseimbangan antara fleksibilitas dan kontrol, yang memungkinkan model yang dihasilkan dapat menangkap pola kompleks serta mempertahankan interpretabilitas dan stabilitas. Oleh karena itu, regresi spline merupakan metode analisis yang cocok digunakan dalam meneliti hubungan antara tingkat partisipasi angkatan kerja dengan persentase kemiskinan, Penelitian dengan judul "Pemodelan Partisipasi Angkatan Kerja (TPAK) terhadap Persentase Penduduk Miskin di Jawa Timur tahun 2023 dengan Metode B-Spline" tertarik pada latar belakang masalah tersebut.

II. METODE PENELITIAN

Variabel penelitian yang digunakan adalah Tingkat Partisipasi Angkatan Kerja (TPAK) dan Persentase Penduduk Miskin di Jawa Timur pada tahun 2023. Variabel tersebut memiliki amatan 38 kab/kota yang bersumber dari *website* BPS Jawa Timur. Tahapan analisis yang digunakan adalah regresi non parametrik *B-spline*. Regresi *B-Spline* merupakan model polinomial tersegmen pada suatu titik fokus yang disebut knot. Hal tersebut memberikan sifat fleksibilitas yang baik (Lyche and Morken). Menurut Budiantara *et al.* (dalam Sugiarti, 2012). Adapun langkah-langkah dari metode regresi non parametrik *B-Spline* adalah sebagai berikut.

1. Eksplorasi Data

Tahapan ini dilakukan untuk melihat pola data dan deskripsi data yang akan dianalisis. Pada regresi *B-spline* eksplorasi data dapat digunakan sebagai salah media dalam mengonfirmasi penggunaan metode. Salah satu metode yang dapat digunakan adalah *scatterplot*. *Scatterplot* digunakan untuk menunjukkan hubungan antara dua variabel kuantitatif yang diukur pada individu yang sama. Setiap individu dalam data direpresentasikan sebagai titik pada grafik, memungkinkan visualisasi pola atau korelasi antara kedua variabel tersebut (Madden et al., 1991). Jika sebaran amatan yang ditampilkan pada *scatterplot* menyebar acak, maka metode regresi *B-Spline* dapat dilanjutkan sebagai metode analisis.

2. Estimasi Parameter

Dalam membentuk model *B-Spline* berorde m , dibutuhkan tambahan knot sebanyak $2m$. β_j dapat diduga dengan *ordinary least square* (Sugiarti, 2012). Adapun model umum regresi nonparametrik *B-Spline* berorde m dengan k -knot u_1, u_1, \dots, u_k dinyatakan sebagai berikut (Sugiarti, 2012) :

$$y_i = \sum_{j=1}^{m+k} \beta_j B_{j-m,m}(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

Keterangan :

β_j : parameter model

$B_{j-m,m}$: merupakan basis *B-Spline* ke- j berorde m , dengan $m = 2, 3, 4, \dots$

x_i : variabel independen ke- i

k : jumlah knots

ε_i : sisaan ke- i

$$B_{j,m}(x) = \frac{x - k_j}{k_{j+m-1} - k_j} B_{j,m-1}(x) + \frac{k_{j+m} - x}{k_{j+m} - k_{j+1}} B_{j+1,m-1}(x) \quad (2)$$

$$B_{j+1}(x) = \begin{cases} 1, & x \in [k_j, k_{j+1}] \\ 0, & \text{untuk } x \text{ yang lainnya} \end{cases} \quad (3)$$

3. Menentukan orde dan titik knot optimal

Orde dan titik knot optimal dapat ditentukan menggunakan metode GCV minimum. GCV merupakan singkatan dari *generalized cross validation* dimana metode ini mampu mengestimasi kesalahan prediksi dalam model serta menentukan orde dan titik knot optimal dalam model. Nilai GCV dapat diduga menggunakan persamaan berikut (Sayuti et al., 2013):

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{i=1}^n \{y_i - f_\lambda(x_i)\}^2}{n\{1 - n^{-1}tr(S_\lambda)\}^2} \quad (4)$$

dengan f_λ adalah estimator dari *smoothing spline* dan $tr(S_\lambda) \leq n$. Nilai λ dipilih dari nilai GCV (λ) yang minimum

4. Akurasi Model

Dalam statistik dan *machine learning*, *Mean Squared Error* (MSE) merupakan ukuran evaluasi yang secara luas digunakan sebagai tolak ukur seberapa akurat suatu model yang diperoleh. MSE menghitung rata-rata kuadrat selisih antara nilai prediksi model dan nilai sebenarnya dari data yang diamati. Bentuk umum dari MSE adalah sebagai berikut (Nuha, 2023) :

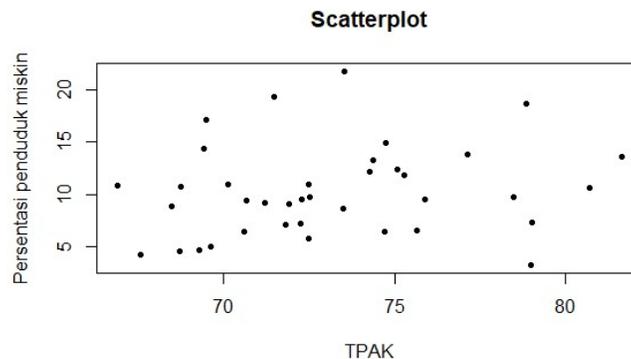
$$MSE = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2 \tag{5}$$

dengan N adalah jumlah sampel data, y_n merupakan nilai sebenarnya dari data untuk n, dan \hat{y}_n adalah nilai prediksi model untuk n.

III. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data

Scatterplot pada Gambar 1 menggambarkan hubungan antara variabel yang dianalisis. Pada Gambar 1 terlihat jika data tersebut tidak membentuk pola tertentu, sehingga kurva regresi yang diperoleh tidak diketahui polanya. Oleh karena itu analisis yang akan digunakan pada penelitian adalah analisis regresi nonparametrik



Gambar 1. Scatterplot TPAK terhadap Persentase Penduduk Miskin.

Statistik deskriptif dari variabel TPAK dan Persentase Penduduk Miskin dapat dilihat pada Tabel 1. Pada Tabel 1 diketahui banyak data pengamatan adalah sebanyak 38, dengan rata-rata dari Persentase Penduduk Miskin yaitu 10,293 dan rata-rata Tingkat Partisipasi Angkatan Kerja yaitu 73,16. Nilai Simpangan baku dari Persentase Penduduk Miskin adalah 4,321289, sedangkan simpangan baku dari Tingkat Partisipasi Angkatan Kerja adalah 3,76715. Nilai minimum dari Persentase Penduduk Miskin adalah 3,310 sedangkan nilai minimum dari Tingkat Partisipasi Angkatan Kerja adalah 66,89. Nilai maksimum dari Persentase Penduduk Miskin adalah 21,760 sedangkan nilai maksimum dari Tingkat Partisipasi Angkatan Kerja adalah 81,64.

Tabel 1. Statistik Deskriptif Variabel (Y) dan Variabel (X)

Variabel	N	Mean	Standar Deviasi	Min	Q ₁	Median	Q ₃	Max
Persentase Penduduk Miskin (Y)	38	10,293	4,321289	3,310	7,188	9,665	12,360	21,760
Tingkat Partisipasi Angkatan Kerja (X)	38	73,16	3,76715	66,89	70,24	72,50	75,23	81,64

3.2 Regresi B-Spline

Sebelum membangun model regresi *B-Spline* perlu diperhatikan penentuan orde, jumlah knot, serta lokasi titik knot. Ketiga hal tersebut dapat ditentukan menggunakan nilai GCV minimum yang terdapat pada tabel 2 berikut.

Tabel 2. GCV dan knot optimal

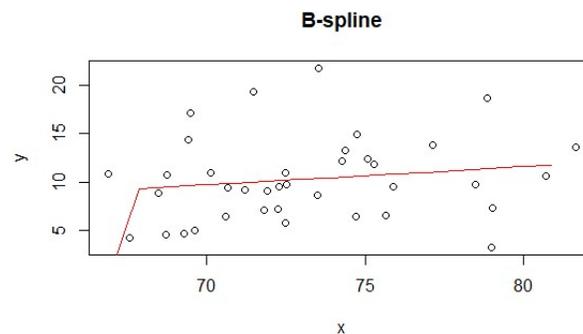
Orde	Jumlah Knot Optimal	Knot Optimal	GCV
2	1	67,89	20,48508
3	1	67,89	21,56097
4	1	67,89	22,81086

Jumlah knot ideal untuk masing-masing orde model adalah 1, menurut Tabel 2. Sebagai model terbaik, yang dipilih adalah model dengan orde sebanyak 2 yang memiliki nilai GCV paling minimum, sehingga dilakukan pendugaan parameter untuk model regresi dengan orde sebanyak 2. Hasil pendugaan parameter disajikan pada Tabel 3.

Tabel 3. Pendugaan parameter

Koefisien	Estimasi	Standar Error	T-Stat	P
bs (x, knots=c(K), degree=1) 1	9,358	1,405	6,660	9.21e-08
bs (x, knots=c(K), degree=1) 2	11,920	1,915	6,224	3.49e-07
F-Statistik : 93,07			p-value : 5.943e-15	
Multiple R ² : 0,8379			Adjusted R ² : 0,8289	

Dari hasil pendugaan parameter pada Tabel 3. Diperoleh nilai F-Statistik sebesar 93,07 dan p-value sebesar 5.943e-15 (lebih kecil dari taraf signifikan) dimana artinya model yang diestimasi secara keseluruhan signifikan dan sangat baik dalam menjelaskan data, Dengan nilai koefisien determinasi sebesar 0,8289, artinya Tingkat Partisipasi Angkatan Kerja (TPAK) mampu menjelaskan model sebesar 82,89% dan sisanya dijelaskan oleh faktor lain diluar model



Gambar 3. Plot Estimasi B-Spline

$$y = 9,358 + 11,920(x - 67,89)^2 + \varepsilon$$

Dari parameter yang diperoleh, ketika persentasi penduduk miskin tidak dipengaruhi oleh tingkat partisipasi angkatan kerja, maka persentasi penduduk miskin mengalami kenaikan sebesar 9,358 per satu satuan. Ketika nilai tingkat partisipasi angkatan kerja berada pada 67,89 atau lebih besar, maka persentasi penduduk miskin mengalami kenaikan sebesar 11,920 per satu satuan

IV. KESIMPULAN

Merujuk hasil penelitian dengan metode regresi nonparametrik *B-Spline* dapat disimpulkan bahwa dengan menggunakan estimasi *B-Spline* akan menghasilkan plot yang halus. Kemudian berdasarkan nilai MSE model *B-Spline* untuk data TPAK di Provinsi Jawa Timur tahun 2023, diperoleh hasil bahwa model regresi *B-Spline* memberikan nilai MSE sebesar 20,11447 sehingga dapat disimpulkan bahwa model *B-Spline* merupakan model yang baik dan cocok digunakan. Dengan model sebagai berikut:

$$y = 9,358 + 11,920(x - 67,89)^2 + \varepsilon$$

Dengan nilai koefisien determinasi sebesar 0,8289, artinya TPAK mampu menjelaskan model sebesar 82,89%. Pada pemodelan regresi non parametrik sangat banyak metode yang bisa digunakan, saran untuk penelitian selanjutnya adalah mengupdate data terbaru pada penelitian ini serta menggunakan beberapa metode yang berbeda untuk dibandingkan sehingga dapat memberikan model yang terbaik kedepannya

DAFTAR PUSTAKA

- BPS (2023). Data Tingkat Partisipasi Angkatan Kerja (TPAK) dan Persentase Penduduk Miskin per Kabupaten/Kota di Jawa Timur Tahun 2023. Jakarta: BPS.
- Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker.
- Geysler, O. F. (2005). Scatterplots and Regression. 1–18. <https://doi.org/10.1002/0471704091.ch1>
- Härdle, W. (1994), *Applied Nonparametric Regression*, Berlin: Springer-Verlag
- Istiqomatul Fajriyah Yulianti, and Sihombing, P. (2020), “Pemodelan Fertilitas Di Indonesia Tahun 2017 Menggunakan Pendekatan Regresi Nonparametrik Kernel dan Spline,” *Jurnal Statistika dan Aplikasinya*, 4, 48–60.
- Madden, M. C., Friedman, M., Keyes, L. L., Koren, H. S., & Burleson, G. R. (1991). Effects of phosgene exposure on lung arachidonic acid metabolism. *Inhalation Toxicology*, 3(1), 73–90. <https://doi.org/10.3109/08958379109145275>
- Nuha, H. (2023). Mean Squared Error (MSE) dan Penggunaannya. *Papers.Ssrn.Com*, 52, 1–1. <https://ssrn.com/abstract=4420880>
- Prima Sukmaraga, 2011. Analisis Pengaruh Indeks Pembangunan Manusia, PDRB Per Kapita, dan jumlah pengangguran Terhadap Jumlah Penduduk Miskin Di Provinsi Jawa Tengah, Universitas Diponegoro.
- Ruppert, D., Sheather, S. J. and Wand, M. P. “An effective bandwidth selector for local least squares regression”. *Journal of the American Statistical Association*, 90, 1257–1270. 1995.
- Sanusi, W., Syam, R., & Adawiyah, R. (2020). Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar). *Journal of Mathematics, Computations, and Statistics*, 2(1), 70. <https://doi.org/10.35580/jmathcos.v2i1.12460>
- Sayuti, A., Kusnandar, D., & Mara, M. N. (2013). Generalized Cross Validation Dalam Regresi Smoothing Spline. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 02(3), 191–196.
- Silverman, B., and Green, P. (1993), *Nonparametrik Regression and Generalized Linear Models (a Roughness Penalty Approach)*, New York: Chapman & Hall.
- Sukirno, S., 2000. *Makroekonomi Modern*. Jakarta: Raja Grafindo Pustaka.
- Sumarsono, Sonny. 2009. *Ekonomi Sumber Daya Manusia Teori dan Kebijakan Publik*. Yogyakarta: Graha Ilmu.
- Tambunan, Dr. Tulus T.H. 2001. *Perekonomian Indonesia Beberapa Masalah Penting*. Jakarta: Ghalia Indonesia.
- World Bank. 2004. *Making The New Indonesia Work for The Poor*. Jakarta: Indopov.