

PEMODELAN ANGKA KEMISKINAN DENGAN REGRESI LINIER SPLINE BERGANDA

Rahmat Hidayat¹, Ma'rufi², Muhammad Ilyas³
Universitas Cokroaminoto Palopo

dayatmath@gmail.com

Abstrak. Keberhasilan pembangunan suatu daerah dapat dilihat dari tingkat kemiskinannya. Dalam mewujudkan tujuan kesejahteraan, pemerintah terus menerus melakukan program pembangunan nasional yaitu pengentasan kemiskinan. Provinsi Jawa Tengah merupakan salah satu provinsi dengan angka kemiskinan cukup tinggi, bahkan melebihi angka kemiskinan nasional. Oleh karena itu, akan diteliti tentang faktor-faktor yang mempengaruhi kemiskinan di Provinsi Jawa Tengah. Metode yang digunakan adalah Regresi *Spline* Berganda karena data yang digunakan memiliki pola yang tidak menentu sehingga tidak tepat jika dianalisis hanya dengan Regresi Linier Berganda. Tujuan dari penelitian ini adalah agar pemerintah setempat dapat menentukan program penanggulangan kemiskinan yang sesuai. Persentase rumah tangga miskin pada setiap kabupaten/kota dijadikan sebagai peubah respon dalam penelitian ini yang dimodelkan dengan faktor yang diduga mempengaruhinya. Data yang digunakan adalah data dari Badan Pusat Statistik. Peubah penjelas digunakan adalah pengeluaran per kapita per orang dalam satu tahun, rata-rata lama sekolah, dan angka harapan hidup. Tahapan analisis diawali dengan melakukan pemotongan kurva berdasarkan fungsi yang bersesuaian. Lalu dipilih model dengan nilai GCV terkecil dari beberapa model dengan *knot* yang berbeda. Hasilnya, model yang terpilih adalah model dengan *knot* tiga yang nilai GCV-nya sebesar 11.078. Peubah penjelas yang berpengaruh nyata terhadap persentase penduduk miskin di Provinsi Jawa Tengah adalah rata-rata lama sekolah.

Kata Kunci: kemiskinan, *knot*, *spline*.

1. Pendahuuan

Regresi nonparametrik yang diawali dengan penelitian tentang Histogram dan Kernel (Hardle, 1990), selanjutnya mendapat perhatian yang sangat luas dari para peneliti, dan terdapat banyak model pendekatan regresi nonparametrik yang telah dikembangkan seperti spline (Hidayat, 2020), Deret Fourier dan Wavelets (Antoniadis, *et al.*, 2001), Deret Ortogonal (Eubank, 1999).

Diantara model-model regresi nonparemetrik yang disebutkan di atas, regresi spline memiliki beberapa keistimewaan yaitu spline merupakan model yang mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik (Budiantara, 2015). Spline bisa memodelkan data pada pola data yang berubah-ubah pada sub-sub interval tertentu karena spline merupakan salah satu jenis potongan polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih daripada polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal pada suatu fungsi atau data.

Spline juga merupakan generalisasi dari ridge regresi. Karena ridge regresi yang menangani multikolonieritas maka dalam spline multikolonieritas tidak perlu lagi dicek dalam data kita. (Wahba, 1990) menunjukkan bahwa spline memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. Dalam analisis regresi spline jika terdapat satu variabel respon dan satu prediktor maka regresi ini disebut regresi spline univariabel. Sebaliknya, jika terdapat satu variabel respon dan lebih dari satu variabel prediktor, maka regresi tersebut dinamakan regresi spline multivariabel (Hidayat, 2019; Chaubey, 2012).

Keberhasilan pembangunan suatu daerah dapat dilihat dari tingkat kemiskinannya. Semakin menurun tingkat kemiskinan suatu daerah dapat diartikan bahwa semakin meningkat kesejahteraan masyarakat daerahnya. Dalam mewujudkan tujuan kesejahteraan, pemerintah terus menerus melakukan program pembangunan nasional yaitu pengentasan kemiskinan. Di Indonesia, persentase kemiskinan mencapai angka 11,13% yang artinya bahwa lebih dari sepersepuluh penduduk Indonesia masih masuk ke dalam kategori miskin. Di Pulau Jawa,

persentase kemiskinan penduduknya masih cukup tinggi, padahal Pulau Jawa merupakan tempat tumbuhnya perekonomian terbesar di Indonesia. Daerah yang memiliki persentase kemiskinan tertinggi di Pulau Jawa adalah Provinsi Jawa Tengah dengan persentase kemiskinan sebesar 13,32% (BPS 2017).

Salah satu metode yang dapat digunakan untuk mendapatkan faktor-faktor yang berpengaruh terhadap kemiskinan diantaranya adalah analisis Regresi Berganda. Namun, metode ini masuk ke dalam metode parametrik, yaitu terdapat asumsi bahwa data harus memenuhi asumsi apabila data yang dimiliki tidak membentuk suatu pola tertentu sehingga tidak memenuhi asumsi bagi analisis dengan metode parametrik. Oleh karena itu, digunakan metode non parametrik yaitu Regresi *Spline* Berganda (Becher, 2009). Penelitian ini bertujuan untuk menelusuri dan mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan di Provinsi Jawa Tengah serta membangun model prediksi dari Regresi Linier *Spline* Berganda.

2. Metode Penelitian

2.1 Analisis Regresi Linier *Spline* Berganda

Spline adalah potongan polinomial yang mulus dan memungkinkan bersifat tersegmen (Herawati 2011). Sedangkan regresi *spline* merupakan regresi yang terdiri atas beberapa penggal polinom berorde tertentu yang saling bersambung pada titik-titik yang disebut dengan *knot* (Smith 1979). Penempatan dan jumlah *knot* serta pemilihan basis akan menentukan kebaikan regresi *spline* (Hardle 1994). Jumlah *knot* perlu ditetapkan terlebih dahulu dan penempatannya dapat dilakukan dengan mencoba semua kombinasi *knot* yang mungkin (Steinberg *et al.* 2001). Menurut Biedermann *et al.* (2009), Fungsi *spline* pada suatu fungsi f dengan orde linier dan sebanyak j peubah bebas dapat dinyatakan sebagai berikut :

$$f(X) = \beta_0 + \sum_{j=1}^p \left[\beta_{(j+(r \times (j-1)))} X_j + \sum_{i=1}^r \beta_{(j+(r \times (j-1))+i)} (X_j - k_{ji})_+ \right]$$

dimana $\beta_0, \beta_1, \dots, \beta_q$ adalah himpunan koefisien atau konstanta real, X_j adalah peubah penjelas ke- j , k_{ji} adalah knot ke- i pada peubah X_j , dan r adalah

banyaknya knot dalam peubah penjelas ke- j . Tanda $+$ pada $(X_j - k_{ji})_+$ menunjukkan bahwa jika hasil dari $(X_j - k_{ji}) \geq 0$ maka nilai tersebut digunakan sebagai hasil, dan jika $(X_j - k_{ji}) < 0$ maka hasil dari perhitungan dianggap nol. Estimasi regresi nonparametrik *spline* untuk penduga β dengan metode *Maximum Likelihood Estimator (MLE)* yang errornya berdistribusi normal akan menghasilkan estimator:

$$f(y; f(X), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y - f(X))^2}{2\sigma^2} \right], f(X) > 0, \sigma^2 > 0$$

Fungsi *likelihood* dapat dituliskan sebagai berikut:

$$\begin{aligned} L(y, f) &= \prod_{a=1}^n f(y_a; f(x_a), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{a=1}^n (y_a - f(x_a))^2 \right] \end{aligned}$$

Dugaan titik fungsi f diperoleh dengan memaksimumkan fungsi *likelihood* $L(y, f)$ yang dapat dirumuskan sebagai berikut :

$$\max_f \{L(y, f)\} = \max_{\beta \in R^{p+r}} \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{a=1}^n \left(y_a - \sum_{j=1}^p \beta_j x_a^j + \sum_{i=1}^r \beta_{(p+i)} (x_a - k_i)_+^p \right)^2 \right) \right\}$$

Menurut Putra *et. al* 2015, apabila optimasi ini diselesaikan akan diperoleh penduga untuk β adalah:

$$\hat{\beta} = (X'X)^{-1} X'y$$

Dengan X adalah matriks peubah penjelas dan y adalah vektor peubah respon.

2.2 B-Spline

Model regresi nonparametrik dengan *B-Spline* merupakan salah satu jenis regresi *spline*. Regresi ini cukup baik digunakan saat data yang dimiliki didekati dengan orde linier. Cara kerja regresi ini adalah dengan mentransformasi nilai peubah penjelas ke dalam nilai basis yang dihitung dengan fungsi rekursif :

$$B_{i,q}(X) = \frac{X - k_i}{k_{i+q-1} - k_i} B_{i,q-1}(X) + \frac{k_{i+q} - X}{k_{i+q} - k_{i+1}} B_{i,q}(X)$$

dengan

$$B_{i,l}(x) = \begin{cases} 1, & X \in [k_i, k_{i+1}] \\ 0, & \text{lainnya} \end{cases}$$

untuk nilai $i = -(m-1), \dots, r$ dan $B_{i,q}$ merupakan nilai basis *B-Spline* (Eberly 1999). Setelah mendapatkan hasil dari nilai basis, maka nilai ini digunakan sebagai peubah penjelas untuk diregresikan terhadap peubah respon. Sehingga fungsi regresinya berubah menjadi :

$$y = B\beta + \varepsilon$$

Penduga dari β akan dihasilkan dari nilai :

$$\hat{\beta} = (B'B)^{-1} B'y$$

Dengan B adalah matriks dari nilai basis (Rodriguez 2001).

2.3 Pemilihan Titik *Knot* Optimal

Penentuan letak *knot* dapat dilakukan dengan beberapa cara, yaitu menurut Huang dan Shen (2004), *knot* dapat ditempatkan secara merata sehingga jarak antara dua *knot* yang berdekatan adalah sama (disebut *knot* berjarak sama). Kemudian *knot* juga dapat ditempatkan secara merata sehingga jumlah amatan dari peubah x sama antara dua *knot* yang berdekatan (disebut *knot* kuantil). Selain itu, penentuan *knot* dapat dilakukan dengan cara seleksi dengan melihat perubahan intersep dari garis regresi (Wu dan Zhang 2006). Pemilihan estimator regresi *spline* terbaik diantara model-model yang didapatkan dilihat berdasarkan kriteria *Generalized Cross Validation* (GCV) yang minimum. Rumus untuk menghitung GCV adalah sebagai berikut:

$$GCV = \frac{MSE}{\left(\frac{1}{n} \text{tr}[I - A(k)] \right)^2}$$

dengan I adalah matriks identitas, k adalah titik *knot* ($k_1, k_2, \dots, k_n$), $A(k) = X(X'X)^{-1}X'$, n adalah banyaknya pengamatan, dan MSE (*Mean Square Error*) merupakan ukuran kinerja estimator (Eubank 1999).

3. Hasil Dan Pembahasan

3.1 Penentuan Titik Knot

Titik *knot* dapat ditentukan dengan beberapa cara yaitu dengan membagi jarak antar *knot* sama besar, membagi jumlah amatan sama banyak (*knot* kuantil), dan menentukan posisi secara eksploratif sesuai dengan perubahan bentuk kurva. Cara ketiga yaitu menentukan posisi *knot* secara eksploratif merupakan cara yang paling efektif karena mengikuti perubahan garis regresi. Namun, untuk jumlah data yang besar akan lebih efektif jika menggunakan cara membagi jarak antar *knot* sama besar atau dengan membagi jumlah amatan sama banyak (*knot* kuantil) karena akan mempercepat proses analisis. Variabel yang diduga mempengaruhi persentase penduduk miskin (y) yakni pengeluaran perkapita (x_1), rata-rata lama sekolah (x_2), dan angka harapan hidup (x_3).

Tabel 1 Posisi *knot* dengan cara eksploratif

Peubah	Knot	Posisi knot		
		1	2	3
x_1	1	10312		
	2	9102	10312	
	3	8700	9533	10793
x_2	1	7,53		
	2	6,33	7,53	
	3	6,41	7,04	8,16
x_3	1	74,51		
	2	73,00	74,51	
	3	73,27	74,27	75,86

Hasil *knot* yang didapatkan dengan cara eksploratif memiliki nilai yang cukup beragam. Hal ini dikarenakan tiap peubah penjelas memiliki satuan yang berbeda. Posisi *knot* yang dihasilkan akan menjadi titik potong dari data yang akan di regresi. Pada pemotongan *knot* satu untuk setiap peubah penjelas, akan didapatkan dua bagian dari garis regresi namun tetap dalam satu model. Garis regresi pertama pada *knot* satu akan meregresikan data pendapatan per kapita per orang di bawah Rp 10 312 dan rata-rata lama sekolah yang kurang dari 7.53 tahun, serta angka harapan hidup di bawah 74.51 tahun. Garis regresi yang kedua pada *knot* pertama akan meregresikan data pendapatan per kapita per orang di atas Rp 10 455 dan rata-rata

lama sekolah yang lebih dari 7.53 tahun, serta angka harapan hidup di atas 74.51 tahun. Setelah kedua garis regresi dijumlahkan, akan terbentuk model Regresi *Spline* Berganda dengan *knot* satu. Begitu juga dengan *knot* dua, akan membagi kurva regresi menjadi tiga bagian dan *knot* tiga akan membagi kurva regresi menjadi empat bagian.

3.2 Pemodelan Regresi *Spline* Berganda

Model yang dipilih dari beberapa pemodelan Regresi *Spline* Berganda adalah model yang memiliki nilai *Generalized Cross Validation* (GCV) yang paling kecil. Hal ini dikarenakan nilai GCV yang kecil merupakan model dengan titik *knot* optimal. Berikut merupakan nilai GCV dari beberapa model dengan *knot* yang berbeda.

Tabel 2 Nilai GCV pada model dengan *knot* berbeda

Jumlah Knot	GCV
1	12,011
2	11,315
3	11,066

Pada Tabel 2 dapat diketahui nilai GCV dari seluruh model yang terbentuk. Model yang paling baik adalah model dengan jumlah *knot* tiga yang memiliki nilai GCV sebesar 11.066. Model regresi dengan *knot* tiga yaitu:

$$\begin{aligned} \hat{y} = & 21,430 - 2,7256x_1 - 0,3869(x_1 - 8700)_+ - 0,9294(x_1 - 9504)_+ + \\ & 2,0490(x_1 - 10793)_+ - 6,7621x_2 - 3,1285(x_2 - 6,41)_+ - \\ & 10,4058(x_2 - 7,04)_+ - 15,3717(x_2 - 8,16)_+ + 0,5011x_3 - \\ & 1,1452(x_3 - 73,27)_+ - 3,7451(x_3 - 74,27)_+ + 2,5161(x_3 - 75,86)_+ \end{aligned}$$

Pada persamaan di atas dapat dilihat model terbaik dari Regresi *Spline* berganda dengan *knot* yang optimum. Model dari persentase penduduk miskin di Provinsi Jawa Tengah ini memiliki nilai koefisien determinasi sebesar 61.5%. Hal ini berarti bahwa keragaman persentase penduduk miskin yang dapat dijelaskan oleh keragaman pendapatan per kapita per orang, rata-rata lama sekolah, dan angka harapan hidup di Provinsi Jawa Tengah adalah sebesar 61.5%. Nilai *Mean Square Error* (*MSE*) dari model tersebut adalah sebesar 11.08. Nilai ini cukup kecil yang berarti bahwa *error* dari nilai prediksi dengan persentase penduduk miskin di

Provinsi Jawa Tengah yang sesungguhnya kecil, sehingga model yang didapatkan sudah cukup baik.

3.3 Pengujian Parameter Regresi *Spline* Berganda

Model yang telah didapatkan dari pemilihan model terbaik dengan *knot* optimum selanjutnya diuji signifikansinya dengan uji serentak. Uji serentak yang digunakan yaitu uji *F*, tujuannya untuk melihat apakah ada pengaruh peubah penjelas terhadap persentase penduduk miskin di Provinsi Jawa Tengah. Setelah diuji, model yang terbentuk memiliki nilai *Fhitung* 2.928 dengan *p-value* sebesar 0.014. Nilai ini bila dibandingkan dengan taraf nyata 0.05 akan menghasilkan kesimpulan tolak H_0 . Hal ini menunjukkan bahwa terdapat peubah penjelas yang berpengaruh nyata terhadap model regresi. Setelah pengujian serentak, selanjutnya dilakukan pengujian parsial dengan uji *t-student*. Tujuan dari pengujian ini adalah untuk melihat peubah apa saja yang memiliki pengaruh nyata terhadap persentase penduduk miskin di Provinsi Jawa Tengah. Nilai dari *p-value* untuk seluruh parameter regresi dapat dilihat dalam Tabel 3

Tabel 3 Nilai dari *p-value* pada uji *t-student*

Prediktor	Parameter	Koefisien	t_{hitung}	<i>p-value</i>
	β_0	21,431	4,537	0.0002*
x_1	β_1	-2,725	-0,675	0.5067
	β_2	-0,387	-0,116	0.9087
	β_3	-0,929	-0,262	0.7955
	β_4	2,049	0,366	0.7178
x_2	β_5	-6,762	-1,741	0.0957
	β_6	-3,128	-0,901	0.3771
	β_7	-10,405	-2,479	0.0213*
	β_8	-15,371	-2,425	0.0240*
x_3	β_9	0,501	0,109	0.9141
	β_{10}	-1,145	-0,263	0.7951
	β_{11}	-3,745	-0,838	0.4111
	β_{12}	2,516	0,423	0.6701

Disajikan nilai p -value dari uji t-student pada model dan terdapat empat parameter regresi yang signifikan terhadap model. Parameter β_0 dengan nilai 21.4309 yang merupakan intersep berpengaruh nyata terhadap model. Parameter β_7 dan β_8 dengan nilai koefisien berturut-turut -10.4058 dan -15.3717 yang keduanya merupakan koefisien dari peubah x_2 juga berpengaruh signifikan terhadap model. Oleh karena itu, dapat dikatakan bahwa peubah penjelas yang berpengaruh nyata terhadap persentase kemiskinan di Provinsi Jawa Tengah adalah rata-rata lama sekolah.

4. Kesimpulan

Berdasarkan hasil yang diperoleh pada bagian sebelumnya, diperoleh bahwa model Spline yang paling baik untuk data persentase kemiskinan di Provinsi Jawa Tengah adalah model Spline dengan tiga titik knot dan peubah penjelas yang berpengaruh nyata terhadap persentase kemiskinan di Provinsi Jawa Tengah adalah rata-rata lama sekolah.

5. Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) yang telah menyediakan data yang digunakan dalam penelitian ini serta semua pihak yang mendukung agar terselesaikannya tulisan ini.

Daftar Pustaka

- Antoniadis, A., Bigot, J. dan Spatinas, T. (2001), "Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study", *Journal of Statistical Software*, 6, pp. 1-83.
- Budiantara, I. N., Ratnasari, V., Ratna, M., dan Zain, I. (2015), "The Combination of Spline and Kernel Estimator for Nonparametric regression and its Properties", *Applied Mathematics Sciences*, 9, pp. 6083-6094.
- Becher, H., Kauermann, G., Khomski, P., dan Kouyate, B. (2009), "Using Penalized Splines to Model Age and Season of Birth Dependent Effects of Childhood Mortality Risk Factors in Rural Burkina Faso", *Biometrical Journal*, 51, pp. 110-122.
- Biedermann S, Dette H, Woods DC. 2009. Optimal Design for Multivariable Spline Models [paper]. Highfield(UK): University of Southampton.

- Chaubey, Y. P., Laib, N., dan Li, J. (2012), “Generalized Kernel Regression Estimator for Dependent Size-Blazed Data”, *Journal of Statistical Planning Inference*, 142, pp. 708-727.
- Eberly D. 1999. *B-Splines Interpolation on Lattices* [internet]. [diunduh 2017 Mei 10]. Tersedia pada : <http://www.e-bookspdf.org>.
- Eubank, R. L., (1999), *Spline Smoothing and Nonparametric Regression*, Marcel Deker: New York.
- Hardle W. 1994. *Applied Nonparametric Regression*. Berlin(DE): Humboldt University.
- Herawati N. 2011. Regresi Spline untuk Pemodelan Bidang Kesehatan: Studi tentang Knot dan Selang Kepercayaan. *Jurnal Ilmu Dasar*, Vol.12: 152- 160. Lampung(ID): Universitas Lampung.
- Hidayat, R., Budiantara, I. N., Otok, B. W., and Ratnasari, V. (2020), “The Regression Curve Estimation by using Mixed smoothing Spline and Kernel (MsS-K) Model”, *Communications in Statistics-Theory and Methods*, <https://doi.org/10.1080/03610926.2019.1710201>.
- Hidayat, R., Budiantara, I. N., Otok, B. W., and Ratnasari, V. 2019. Kernel-Spline Estimation of Additive Nonparametric Regression Model. The 9th Basic Science International Conference (BaSIC), March 2019, Malang, Indonesia.
- Huang JZ, Shen H. 2004. Functional Coefficient Regression Models for NonLinear Time Series: A Polynomial Spline Approach. *Scandinavian Journal of Statistics*, Vol. 31: 515–534. Oxford(UK): Blackwell Publishing Ltd.
- Putra IMB, Srinadi IGAM, Sumarjaya IW. 2015. Pemodelan Regresi Spline (Studi Kasus: Herpindo Jaya Cabang Ngaliyan). *Jurnal Matematika*, Vol. 4: 110-114. Bali(ID): Universitas Udayana.
- Rondriguez G. 2001. Smoothing and Non-Parametric Regression [paper]. New Jersey(US): Princeton University.
- Smith PL. 1979. Splines as a Useful and Convenient Statistical Tool. *The American Statistician* Vol. 33: 57-62.
- Steinberg D, Colla PL, Kerry M. 2001. *MARS User Guide*. California(US): Salford System.
- Wahba, G., (1990), *Spline Models for Observation Data*, Pennsylvania, SIAM.
- Wu H, Zhang JT. 2006. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. New Jersey(US): John Wiley and Sons.