

**PEMODELAN REGRESI NONPARAMETRIK *SPLINE*
TRUNCATED DAN APLIKASINYA PADA INDEKS
PEMBANGUNAN MANUSIA DI PULAU KALIMANTAN**

Izzatul Yazidah Al-Fajri Polanagau^{1*}, Sifriyani¹, Wasono¹

¹Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas
Mulawarman, Indonesia

Corresponding author: izzatulyazidah198@gmail.com

Abstrak Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM dapat menentukan peringkat atau level pembangunan suatu wilayah dan Negara. Dalam pembangunan, IPM yang tinggi merupakan sasaran yang diharapkan dapat tercapai terutama bagi negara berkembang. Kalimantan merupakan salah satu wilayah di Indonesia, yang mana pada tahun 2017 nilai IPM mencapai 70,13% yang masih lebih kecil jika dibandingkan dengan IPM nasional di tahun yang sama yaitu sebesar 70,81%. Banyak faktor yang mempengaruhi IPM di pulau Kalimantan sehingga perlu dilakukan pemodelan untuk mengetahui faktor-faktor apa saja yang secara signifikan mempengaruhi IPM. Penelitian ini menggunakan 5 faktor yang diduga mempengaruhi IPM yaitu tenaga kesehatan (x_1), sarana kesehatan (x_2), pertumbuhan ekonomi (x_3), Angka Partisipasi Sekolah SMA (x_4) dan Tingkat Partisipasi Angkatan Kerja (x_5). Data tersebut merupakan data tahun 2017 yang diperoleh dari Badan Pusat Statistik (BPS) di 5 Provinsi di Kalimantan. Metode yang digunakan untuk memodelkan IPM adalah regresi nonparametrik *spline truncated* dengan titik knot optimal yaitu tiga titik knot yang memiliki nilai GCV terkecil. Hasil penelitian diperoleh nilai R^2 sebesar 85,70% yang menunjukkan bahwa model yang terbentuk layak digunakan untuk memodelkan pola data.

Kata Kunci: Regresi nonparametrik, *Spline truncated*, IPM, Knot, GCV

1 PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan ukuran yang digunakan untuk mengevaluasi pembangunan manusia di suatu daerah. IPM yang tinggi dapat meningkatkan pembangunan manusia di daerah tersebut sehingga daerah tersebut dapat berkembang dan menjadi lebih baik dari sebelumnya. Dalam hal ini yang dapat dilakukan untuk meningkatkan nilai IPM di Kalimantan ialah dengan cara mengetahui faktor-faktor yang apa saja yang mempengaruhi IPM di Kalimantan.

Indeks Pembangunan Manusia (IPM) pertama kali diperkenalkan pada tahun 1990 oleh *United Nations Development Programme* (UNDP). Data IPM pada Badan Pusat Statistik (BPS) di Pulau Kalimantan terus mengalami peningkatan setiap tahunnya. Pada tahun tahun 2013 nilai IPM di Kalimantan sebesar 68,02%, tahun 2014 nilai IPM sebesar 68,55%, tahun 2015 sebesar 69,09%, tahun 2016 sebesar 69,57% dan tahun 2017 sebesar 70,13%. Nilai-nilai tersebut merupakan nilai IPM yang ada di 5 Provinsi yaitu Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat dan Kalimantan Utara. Dari nilai tersebut kita dapat melihat bahwa nilai IPM dari tahun 2013-2017 di Pulau Kalimantan terus mengalami peningkatan setiap tahunnya. Meski mengalami peningkatan tapi jika dilihat dari kategori nilai IPM di seluruh provinsi yang ada di Indonesia nilai IPM di Kalimantan masih berada di bawah nilai IPM nasional.

Berdasarkan uraian diatas maka akan dilakukan penelitian untuk mengetahui faktor-faktor apa saja yang mempengaruhi IPM di Kalimantan. Faktor-faktor tersebut dapat diketahui dengan menggunakan metode pemodelan. Metode yang digunakan untuk menyelesaikan masalah pemodelan dalam penelitian ini adalah regresi nonparametrik *spline truncated*.

2 TINJAUAN PUSTAKA

2.1 Regresi Nonparametrik

Regresi nonparametrik merupakan metode dalam analisis regresi yang digunakan ketika kurva regresi antara variabel respon dan variabel prediktor tidak diketahui bentuknya atau polanya. Secara umum model regresi nonparametrik adalah sebagai berikut:

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

dimana y_i merupakan variabel respon, x_i adalah variabel prediktor dan $f(x_i)$ merupakan fungsi regresi yang tidak membentuk suatu pola. Salah satu metode yang ada dalam regresi nonparametrik ialah metode regresi nonparametrik *spline truncated*. Secara umum fungsi *spline* berorde m adalah sembarang fungsi yang dapat di tulis dalam bentuk

$$f(x_i) = \sum_{k=0}^m \beta_k x_i^k + \sum_{h=1}^r \beta_{m+h} (x_i - k_h)_+^m \quad (2)$$

dimana $\sum_{k=0}^m \beta_k x_i^k$ merupakan komponen polinomial dan $\sum_{h=1}^r \beta_{m+h} (x_i - k_h)_+^m$ merupakan komponen *truncated*.

2.2 Pemilihan Titik Knot Optimal

Terdapat beberapa metode yang dapat digunakan untuk memilih titik *knot* optimal, salah satunya adalah dengan menggunakan metode GCV (*Generalized Cross Validation*). Berikut merupakan rumus untuk mencari nilai GCV [3]

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{(n^{-1} \text{trace}[I - A((K_1, K_2, \dots, K_r))])^2} \quad (3)$$

dengan

$$MSE(K_1, K_2, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (4)$$

K_1, K_2, \dots, K_r adalah titik knot dan matriks $A(K_1, K_2, K_r)$ diperoleh dari persamaan $\hat{y} = A(K_1, K_2, \dots, K_r)_y$.

2.3 Pengujian Signifikansi Parameter Model Regresi

Terdapat dua tahap pengujian parameter yaitu pengujian secara simultan dan pengujian secara parsial.

2.3.1 Uji Simultan

Uji simultan merupakan pengujian parameter model regresi secara bersamaan. Uji ini bertujuan untuk mengetahui parameter model regresi tersebut telah signifikan atau tidak. Rumusan hipotesis untuk pengujian signifikansi simultan adalah [4]

Hipotesis

$$H_1 : \beta_1 = \beta_2 = \dots = \beta_{21} = 0$$

$$H_1 = \text{minimal terdapat satu } \beta_k \neq 0, k = 1, 2, \dots, 21$$

Statistik uji untuk pengujian signifikansi simultan adalah [4]

$$F_{hitung} = \frac{MS_{regresi}}{MS_{error}} \quad (5)$$

Daerah penolakan : tolak H_0 jika $F_{hitung} \geq F_{\alpha(m-1, n-m)}$. Atau $p\text{-value} < \alpha$.

2.3.2 Uji Parsial

Pengujian parameter secara parsial bertujuan untuk mengetahui apakah variabel-variabel prediktor berpengaruh secara parsial terhadap variabel respon. Rumusan hipotesis untuk pengujian signifikansi parsial adalah [4]

Hipotesis

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Statistik uji untuk pengujian signifikansi parsial adalah (Merdekawati dan Budiantara, 2013)

$$t_{hitung} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \quad (6)$$

Daerah penolakan : H_0 ditolak apabila $|t_{hitung}| > t_{((\alpha/2), (n-m))}$.

2.4 Literatur Review

Penelitian dibidang Regresi nonparametrik spline terus dikembangkan oleh para ahli dibidang statistika, diantaranya menentukan estimasi model GWR dengan pendekatan *spline* [5]. Kajian teoritis yang diuraikan dalam penelitian ini didapatkan hasil estimator model GWR dengan pendekatan *spline* dan sifat estimator model yang unbiased dan linear terhadap observasi. Penelitian selanjutnya pengembangan model *nonparametric geographically weighted regression* dengan menggunakan pendekatan *spline truncated* [6]. Penelitian ini membahas tentang prosedur estimasi MLE terboboti pada model *nonparametric geographically weighted regression* dengan menggunakan pendekatan *spline truncated* dan menyelidiki sifat-sifat estimator model.

3 SUMBER DATA DAN METODOLOGI

3.1 Sumber Data

Data pada penelitian ini merupakan data sekunder yang diambil pada website resmi Badan Pusat Statistik (BPS) Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat dan Kalimantan Utara pada tahun 2017. Adapun variabel-variabel pada penelitian terdapat sebanyak 1 variabel respon dan 5 variabel prediktor. Variabel responnya ialah Indeks Pembangunan Manusia dan variabel prediktornya ialah tenaga kesehatan, sarana kesehatan, pertumbuhan ekonomi, Angka Partisipasi Sekolah SMA dan Tingkat Partisipasi Angkatan Kerja[2]

3.2 Metodologi

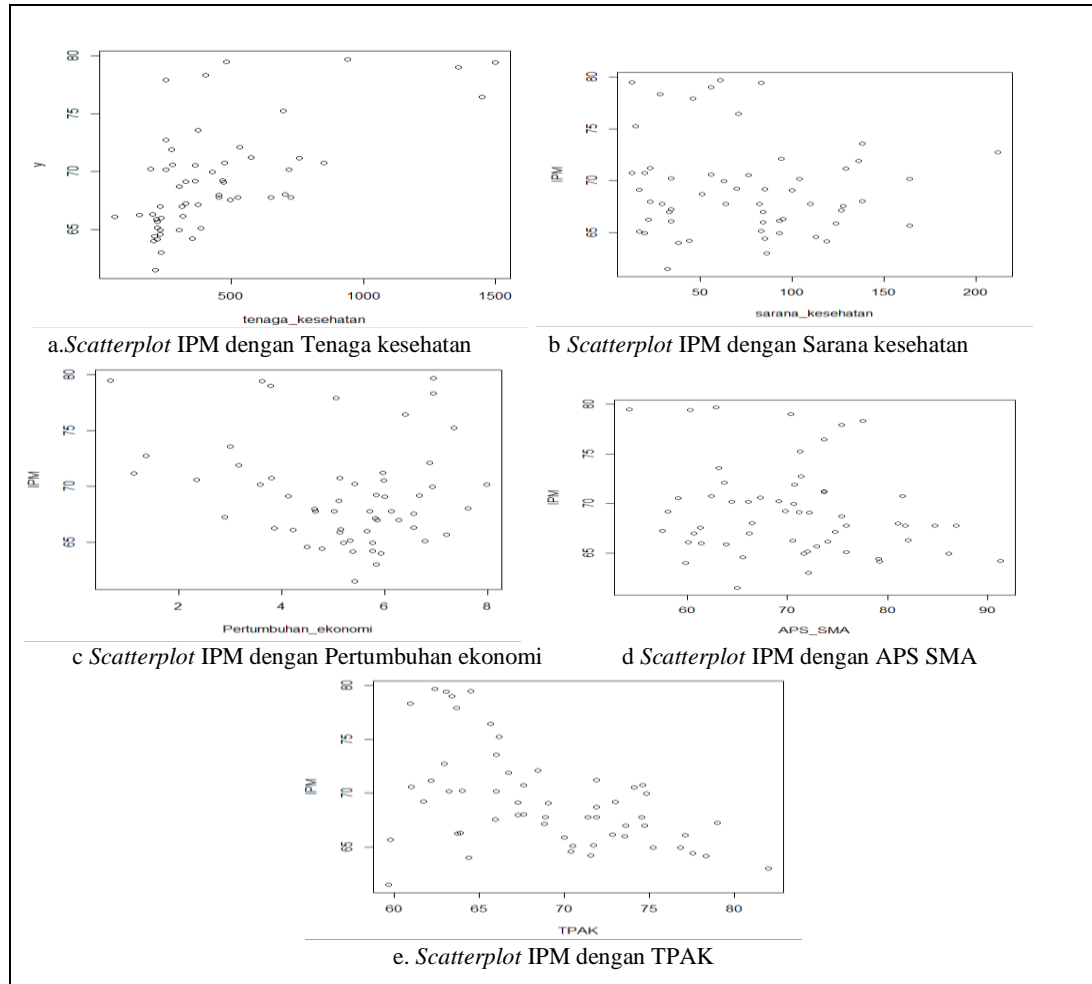
Metode yang digunakan pada penelitian ini ialah metode regresi nonparametrik *spline truncated*. Metode ini terdiri dari 4 tahap yaitu pemilihan titik knot optimal, estimasi model regresi nonparametrik *spline truncated*, uji signifikansi parameter simlutan dan uji signifikansi parameter parsial

4 HASIL DAN PEMBAHASAN

Dalam penelitian ini diduga terdapat 5 variabel yang mempengaruhi IPM. Analisis ini mencakup 56 kab/kota di Kalimantan.

4.1 Pola Hubungan

Sebelum dianalisis menggunakan analisis regresi *spline*, terlebih dahulu diselidiki pola hubungan antara masing-masing variabel prediktor dengan variabel respon dengan menggunakan *scatter plot*.



Gambar 1: Scatterplot antara IPM dan variabel predikto

Berdasarkan gambar 1 diatas dapat dilihat bahwa pola hubungan yang terbentuk antara IPM dengan masing-masing variabel prediktor tidak membentuk suatu pola tertentu oleh sebab itu dapat digunakan metode regresi *nonparametrik spline truncated*.

4.2 Titik Knot Optimal

Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot dan tiga titik knot adalah sebagai berikut

Tabel 1: Titik Knot Optimal

Jumlah Knot	x1	x2	x3	x4	x5	GCV
Satu knot	157,45	26,72	1,18	56,78	61,23	8,65

Dua knot	58 107,72	13 19,86	0,68 0,93	54,22 55,50	59,69 60,46	8,37
Jumlah Knot	x1	x2	x3	x4	x5	GCV
Tiga knot	306,62; 803,86; 1052,48	47,31; 115,93; 150,24	1,94; 4,46; 5,72	60,61; 73,40; 79,80	63,54; 71,25; 75,09	7,29*

Berdasarkan tabel diatas dapat dilihat bahwa nilai GCV yang paling minimum adalah nilai GCV dengan tiga titik knot. Setelah didapatkan nilai titik knot yang optimal langkah berikutnya ialah membuat model regresi nonparametrik yang terbentuk dengan tiga titik knot.

$$y = 0,026x_1 - 0,026(x_1 - 306,62)_+ + 0,042(x_1 - 803,86)_+ - 0,060(x_1 - 1052,48)_+ + 0,125x_2 - 0,152(x_2 - 47,31)_+ - 8,179x_3 + 11,792(x_3 - 1,94)_+ - 6,354(x_3 - 4,46)_+ - 6,354x_3(x_3 - 5,72)_+ + 0,893(x_4 - 73,40)_+ + 1,266x_5 - 1,910(x_5 - 63,54) - 0,980(x_5 - 71,25)_+$$

Model regresi nonparametrik *spline truncated* dengan titik knot ini memiliki nilai R^2 sebesar 85,78% hal ini dapat dikatakan bahwa model ini dapat menjelaskan IPM sebesar 85,78%.

4.1 Pengujian Signifikansi Secara Simultan

Uji serentak dilakukan pada parameter model regresi terhadap variabel IPM secara simultan. Berikut ini merupakan hasil uji simultan model regresi *nonparametric spline*

Tabel 2: Hasil Uji Simultan

Sumber variasi	Derajat Bebas	Jumlah Kuadrat	Rataan Kuadrat	F _{hitung}
Regresi	20	962,08	48,10	10,56
Residual	35	159,50	4,56	
Total	55	1121,58		

Berdasarkan tabel 2 dapat dilihat bahwa nilai F_{hitung} sebesar 4,68. Pada pengujian ini, taraf signifikansi yang digunakan sebesar 5% dan $F_{0,05(20,35)} = 1,88$. Berdasarkan nilai F_{hitung} dan $F_{0,05(20,35)}$ maka dapat diputuskan tolak H_0 karena $F_{hitung}(10,56) > F_{0,05(20,35)}(1,88)$. Sehingga dapat disimpulkan bahwa di dalam model ini terdapat parameter yang signifikan.

4.2 Pengujian Signifikansi Secara Parsial

Uji parsial dilakukan untuk mengetahui parameter mana yang berpengaruh signifikan terhadap model regresi. Berikut merupakan hasil pengujian signifikansi parameter model secara parsial.

Tabel 3: Uji Parsial Estimasi Parameter Spline linier dengan 3 titik knot

Titik knot ke-i	Variabel	Parameter	Nilai Estimasi	P-value	Keputusan
306,6207		β_1	0.0261	0,0022	Signifikan
803,8621		β_2	-0.0263	0,0096	Signifikan

Titik knot ke-i	Variabel	Parameter	Nilai Estimasi	P-value	Keputusan
1052,4828	x_1	β_3	0.0423	0,0360	Signifikan
		β_4	-0.0604	0,0345	Signifikan
47,3103 115,9310 150,2414	x_2	β_5	0.1247	0,0113	Signifikan
		β_6	-0.1516	0,0223	Signifikan
		β_7	-0.0480	0,5452	Tidak signifikan
		β_8	0.0996	0,3379	Tidak signifikan
1,9403 4,4610	x_3	β_9	-8.1789	0,0075	Signifikan
		β_{10}	11.7917	0,0041	Signifikan
		β_{11}	-6.3536	0,0049	Signifikan
		β_{12}	4.7333	0,0070	Signifikan
60,6148 73,4044 79,7993	x_4	β_{13}	0.1412	0,6851	Tidak signifikan
		β_{14}	-0.0731	0,8526	Tidak signifikan
		β_{15}	0.8926	0,0038	Signifikan
		β_{16}	-0.8473	0,0646	Tidak signifikan
63,5417 71,2452 75,0969	x_5	β_{17}	1.2659	0,0039	Signifikan
		β_{18}	-1.9098	0,0006	Signifikan
		β_{19}	0.9796	0,0458	Signifikan
		β_{20}	-0.7349	0,2012	Tidak Signifikan

Berdasarkan tabel 4.2 dapat dilihat bahwa terdapat 21 parameter yang terbentuk dengan taraf signifikansi sebesar 5%. Dari 21 parameter yang terbentuk terdapat 14 parameter yang signifikan dan 7 parameter yang tidak signifikan. Parameter tersebut tidak signifikan karena nilai p -value nya $< \alpha = 0,05$

5 KESIMPULAN

1. Model regresi nonparametrik *spline truncated* terbaik adalah dengan tiga titik knot. Model ini memiliki nilai GCV paling minimum yaitu sebesar 7,2897 dengan R^2 sebesar 85,78%. Berikut adalah model *spline* terbaik

$$y = 0,026x_1 - 0,026(x_1 - 306,62)_+ + 0,042(x_1 - 803,86)_+ - 0,060(x_1 - 1052,48)_+ + 0,125x_2 - 0,152(x_2 - 47,31)_+ - 8,179x_3 + 11,792(x_3 - 1,94)_+ - 6,354(x_3 - 4,46)_+ - 6,354x_3(x_3 - 5,72)_+ + 0,893(x_4 - 73,40)_+ + 1,266x_5 - 1,910(x_5 - 63,54) - 0,980(x_5 - 71,25)_+$$

2. Pada uji simultan secara keseluruhan semua variabel 136redictor berpengaruh terhadap variabel respon dan pada uji individu dari 21 parameter yang terbentuk terdapat 14 parameter yang signifikan dan 7 parameter yang tidak signifikan. Parameter tersebut tidak signifikan karena nilai p -value nya $< \alpha = 0,05$
3. Faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia di kabupaten/kota di Pulau Kalimantan pada tahun 2017 ialah tenaga kesehatan (x_1), sarana kesehatan (x_2), pertumbuhan ekonomi (x_3), Angka Partisipasi Sekolah SMA (x_4) dan Tingkat Partisipasi Angkatan Kerja (x_5).

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik (2017). *Indeks Pembangunan Manusia 2017*. Jakarta: CV Nario Sari
- [2] Badan Pusat Statistik (2017). *Statistik Kesejahteraan Rakyat Provinsi Kalimantan Timur*. Samarinda: CV Mahendra
- [3] Bintaningrum, M.F., & Budiantara I.N. (2014). Pemodelan *Regresi Nonparametrik Spline Truncated* dan Aplikasinya pada Angka Kelahiran Kasar di Surabaya. *Jurnal Sains dan Seni Pomits (Nomor 1 Vol 3)*. Surabaya: Institut Teknologi Sepuluh November.
- [4] Merdekawati, I.P., & Budiantara, I.N. (2013). Pemodelan Regresi *Spline Truncated* Multivariabel pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah. *Jurnal dan Seni Pomits (Nomor 1 Vol 2)*. Surabaya: Institut Teknologi Sepuluh November.
- [5] Sifriyani, Haryatmi, I.N Budiantara, and Gunardi. 2017. Geographically Weighted Regression with Spline Approach. *Far East Journal of Mathematical Sciences*. 101(6). 1183-1196. Scopus
- [6] Sifriyani, S. H. Kartiko, I. N. Budiantara and Gunardi. 2018 Development of Nonparametric Geographically Weighted Regression Using Truncated Spline Approach. *Songklanakarin Journal of Science and Technology*. 40(4). 909-920. Scopus