



Pemodelan Regresi Nonparametrik Spline dengan Pemilihan Titik Knot Optimal Menggunakan Metode *Unbiased Risk* dan *Generalized Cross Validation* pada Kasus Indeks Pembangunan Manusia di Jawa Tengah

Spline Nonparametric Regression Modeling with Optimal Knot Point Selection Using the Unbiased Risk and Generalized Cross Validation Methods in the Case of Human Development Index in Central Java

Winda Nurmalisa^{*}, Indah Manfaati Nur

Program Studi Statistika, Universitas Muhammadiyah Semarang, Semarang

Corresponding author: wnurmalisa@gmail.com*, indahmnur@unimus.ac.id

Riwayat Artikel: Dikirim; Diterima; Diterbitkan

Abstrak

Pada tahun 2017, Jawa Tengah memiliki nilai Indeks Pembangunan Manusia (IPM) sebesar 70,52. Jawa Tengah dengan provinsi yang paling padat penduduknya di pulau Jawa memiliki nilai IPM yang rendah dibandingkan provinsi lain yang ada di pulau Jawa. Hal ini menunjukkan bahwa perkembangan pembangunan di Provinsi Jawa Tengah masih tertinggal dibandingkan provinsi lain yang ada di pulau Jawa. Langkah yang dapat dilakukan pemerintah untuk meningkatkan nilai IPM di Provinsi Jawa Tengah adalah dengan mengetahui faktor apa saja yang berpengaruh terhadap IPM. Tujuan penelitian ini adalah untuk memodelkan IPM dengan regresi nonparametrik spline. Model regresi spline terbaik dihasilkan dari pemilihan titik knot yang optimal. Metode pemilihan titik knot optimal pada penelitian ini, menggunakan *Unbiased Risk* (UBR), *Cross Validation* (CV), dan *Generalized Cross Validation* (GCV). Kriteria pemilihan model terbaik adalah dengan cara membandingkan nilai MSE dan R-square dari ketiga metode yang digunakan, Model terbaik yang didapat dari membandingkan ketiga metode pemilihan titik knot optimal adalah metode *Generalized Cross Validation* (GCV) dengan 3 titik knot. Model tersebut memiliki R-square sebesar 94,32 % dan MSE 0,10. Hal ini membuktikan bahwa pemodelan regresi nonparametrik spline dengan menggunakan metode GCV lebih baik untuk data Indeks Pembangunan Manusia di Jawa Tengah dibandingkan metode CV dan UBR.

Kata Kunci : *Indeks Pembangunan Manusia, Regresi nonparametrik, Spline, GCV, CV, UBR*

Abstract

In 2017, Central Java has a Human Development Index (HDI) value of 70.52. Central Java with the most populous province on the island of Java has a lower HDI value than other provinces on the island of Java. This shows that human development in Central Java Province is still lagging behind compared to other provinces in Java. The step that the government can take to increase the HDI value in Central Java Province is to find out what factors cause the HDI. The purpose of this study is to model HDI with spline nonparametric regression. The best spline regression model results from the selection of optimal knot points. The optimal knot point selection method in this study, using Unbiased Risk (UBR), Cross Validation (CV), and Generalized Cross Validation (GCV). The criterion for selecting the best model is comparing the MSE and R-square values of the three methods used, the best model obtained from the comparison of the three optimal comparison point selection methods is the Generalized Cross Validation (GCV) method with 3 knots. This model has an R-square of 94.32% and MSE 0.10. This proves that the nonparametric spline regression modeling using the GCV method is better for the Human Development Index data in Central Java for the CV and UBR methods.

Keywords: *Human Development Index, Nonparametric Regression, Spline, GCV, CV, UBR*

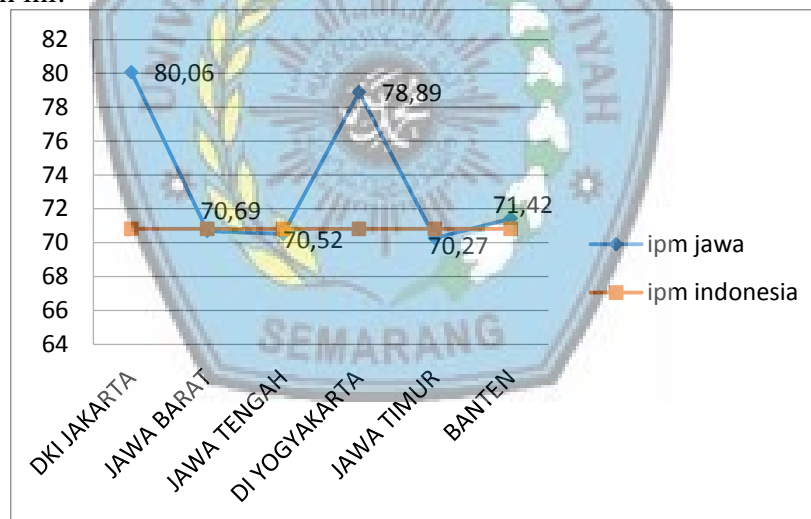


PENDAHULUAN

Manusia merupakan kekayaan bangsa yang sesungguhnya. Oleh karena itu, manusia selalu menjadi sasaran dari pembangunan suatu bangsa. Tujuan utama pembangunan adalah menciptakan lingkungan yang memungkinkan rakyat menikmati umur panjang, sehat dan menjalankan kehidupan yang produktif. Pembangunan manusia menempatkan manusia sebagai tujuan akhir dari pembangunan bukan alat dari dari pembangunan. Keberhasilan pembangunan manusia dapat dilihat dari seberapa besar permasalahan mendasar masyarakat dapat teratasi.

Indeks Pembangunan Manusia menjadi indikator penting untuk mengukur salah satu aspek penting yang berkaitan dengan kualitas dari hasil pembangunan ekonomi, yakni derajat perkembangan manusia. IPM dibangun melalui pendekatan tiga dimensi dasar, dimensi tersebut ialah kesehatan, Pendidikan dan Ekonomi (Melliana, 2013).

Indeks Pembangunan Manusia merupakan hubungan antara manusia dengan pembangunan yang ada disekitarnya, yang mana saling mempengaruhi satu sama lain. Dalam kata lain, terdapat suatu korelasi positif antara nilai IPM dengan derajat keberhasilan pembangunan ekonomi (Tambunan, 2003). Pulau Jawa merupakan pulau dengan jumlah penduduk yang paling padat, salah satu provinsi yang paling padat penduduk nya adalah Jawa Tengah, dengan semakin padat penduduk dalam satu wilayah maka angka IPM harus diperhitungkan. Jawa Tengah dengan provinsi yang padat penduduk nya di Pulau Jawa memiliki nilai IPM pada tahun 2017 sebesar 70.52 % namun nilai IPM tersebut masih rendah jika dibandingkan dengan provinsi lain yang ada di Pulau Jawa. Hal ini dapat dilihat pada gambar 1 dibawah ini:



Gambar 1 Perbandingan Indeks Pembangunan Manusia

Pada gambar 1 diatas terlihat bahwa nilai IPM di provinsi Jawa Tengah masih berada dibawah provinsi lain yang ada dipulau jawa kecuali Jawa Timur dan juga masih berada di bawah nilai IPM Indonesia. Oleh karena itu perlu diketahui faktor-faktor apa saja yang mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa tengah sehingga pemerintahan Provinsi Jawa Tengah bisa lebih memperhatikan dan mengupayakan program pembangunan manusia untuk meningkatkan angka Indeks Pembangunan Manusia sebagai upaya melakukan perbaikan kesejahteraan manusia di Provinsi Jawa Tengah.

Perhitungan IPM didasarkan pada tiga dimensi diantaranya dimensi kesehatan, pendidikan dan standar hidup layak. Dimensi kesehatan diwakili oleh faktor Jumlah fasilitas kesehatan



yang ada di Provinsi Jawa Tengah, dimensi pendidikan diwakili oleh faktor Angka Melek Huruf. Untuk dimensi standar hidup layak diwakili oleh faktor PDRB di provinsi Jawa Tengah, dan faktor selanjutnya yang mempengaruhi indeks pembangunan manusia di Jawa Tengah adalah tingkat pengangguran terbuka.

Salah satu metode yang digunakan untuk mengetahui faktor-faktor yang berpengaruh terhadap IPM di Jawa Tengah adalah analisis regresi. Disini peneliti menggunakan regresi nonparametrik *Spline*. Penggunaan metode regresi nonparametrik *Spline* memiliki kelebihan dikarenakan *Spline* merupakan model yang mempunyai interpretasi statistik dan visual yang sangat khusus serta sangat baik (Budiantara, 2006). Beberapa hal yang perlu diperhatikan dalam membentuk model pada regresi *spline* adalah menentukan orde model, banyaknya titik knot dan lokasi titik knot tersebut (Montoyo, Ulloa dan Miller, 2014). ada beberapa metode yang dapat digunakan untuk memilih titik knot optimal dalam estimator *spline*, antara lain metode *Cross Validation* (CV) (Craven dan Wahba, 1979), *Unbiased Risk* (UBR) (Wahba, 1990; Wang, 1998), *Generalized Maximum Likelihood* (GML) (Eubank, 1999) dan *Generalized Cross Validation* (GCV) (Wahba, 1990).

GCV dan UBR adalah metode pemilihan titik knot optimal yang mempunyai banyak kelebihan. Adapapun kelebihan yang dimiliki metode GCV antara lain adalah sederhana dan efisien dalam perhitungan, optimal secara asimtotik, invarian terhadap transformasi dan tidak memerlukan informasi terhadap varians. Pemilihan titik knot dengan metode GCV akan lebih baik jika digunakan pada data yang *Gaussian* (Berdistribusi Normal), sementara pemilihan titik knot dengan metode UBR akan lebih baik jika digunakan pada data *non-Gaussian* atau tidak berdistribusi normal, (Wahba, 1990; Wang, 1998). Berdasarkan pernyataan tersebut, peneliti tertarik untuk membandingkan apakah model regresi nonparametrik *spline* multivariabel menggunakan metode GCV akan lebih baik jika dibandingkan dengan model regresi nonparametrik *spline* menggunakan metode UBR pada data Indeks Pembangunan Manusia di Provinsi Jawa Tengah pada tahun 2017.

Penelitian mengenai Indeks Pembangunan Manusia dan Analisis dengan menggunakan metode regresi Nonparametrik *Spline* sudah pernah dilakukan sebelumnya, beberapa diantaranya Penelitian Melliana (2013) menggunakan metode regresi data panel yang menghasilkan kesimpulan yaitu IPM di Jawa Timur dipengaruhi oleh rasio terhadap guru, angka partisipasi SMP/MTS, jumlah sarana kesehatan, rumah tangga dengan akses air bersih, kepadatan penduduk, tingkat partisipasi angkatan kerja dan PDRB per kapita. Penelitian Awal (2014) yang menghasilkan faktor-faktor yang berpengaruh terhadap IPM di provinsi Papua menggunakan metode nonparametrik. Penelitian oleh Retno (2014) yang menghasilkan faktor-faktor yang berpengaruh terhadap IPM di Jawa Timur yaitu angka kematian bayi, pertumbuhan ekonomi, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja menggunakan regresi semiparametrik. Penelitian oleh Nisa, I dan Karim, A (2017) yang menghasilkan bahwa indeks pembangunan manusia di Jawa Tengah dipengaruhi oleh pengeluaran perkapita. Pada penelitian ini peneliti ingin melakukan penelitian tentang pemodelan IPM di Jawa Tengah dengan menggunakan pendekatan Regresi Nonparametrik *Spline*.

Metode *spline truncated* dipilih karena *scatter plot* antara variabel respon dengan masing-masing variabel prediktor tidak mengikuti pola tertentu dan Penelitian ini mengkaji GCV dan UBR untuk pemilihan titik knot pada data multivariabel yang diharapkan akan menjadi suatu kajian yang penting bagi pengembangan ilmu statistika. Hasil dari penelitian ini diharapkan mampu memberi informasi serta masukan positif bagi pemerintah Provinsi Jawa Tengah. Sehubungan dengan uraian diatas penelitian ini diberi judul



“Pemodelan Regresi Nonparametrik Spline dengan Pemilihan Titik Knot Optimal Menggunakan Metode *Unbiased Risk* dan *Generalized Cross Validation* pada Kasus Indeks Pembangunan Manusia di Jawa Tengah”.

METODE

1. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder tahun 2017 yang diperoleh dari Badan Pusat Statistik Unit observasi yang digunakan dalam penelitian ini adalah 35 kabupaten/kota di Provinsi Jawa Tengah.

2. Variabel Penelitian

Adapun variabel penelitian yang digunakan dalam penelitian ini terdiri dari variabel respon (Y) Indeks pembangunan Manusia dan variabel prediktor x_1 Angka Melek Huruf, X_2 Jumlah Fasilitas Kesehatan, X_3 PDRB dan X_4 Tingkat Pengangguran Terbuka

3. Langkah Analisis

- 1) Membuat statistik deskriptif dari variabel respon dan variabel prediktor untuk mengetahui karakteristik masing-masing Kabupaten/Kota di Jawa Tengah.
- 2) Membuat *scatterplot* antara variabel prediktor dengan variabel respon untuk mengetahui perilaku pola data.
- 3) Memodelkan Indeks Pembangunan Manusia di Jawa Tengah dengan *Spline* linear 1 knot, 2 knot, dan 3 knot dengan metode *Generalized Cross Validation* (GCV) untuk menentukan titik knot optimal
- 4) Memodelkan Indeks Pembangunan Manusia di Jawa Tengah dengan *Spline* linear 1 knot, 2 knot, dan 3 knot dengan metode *Cross Validation* (CV) untuk menentukan titik knot optimal.
- 5) Memodelkan Indeks Pembangunan Manusia di Jawa Tengah dengan *Spline* linear 1 knot, 2 knot, dan 3 knot dengan metode *Unbiased Risk* (UBR) untuk menentukan titik knot optimal.
- 6) Memilih model spline terbaik berdasarkan nilai R^2 terbesar dan MSE terkecil dari ketiga metode yang digunakan.
- 7) Mendapatkan estimasi parameter dari titik knot optimal berdasarkan metode yang terpilih.
- 8) Melakukan interpretasi model terbaik dan menarik kesimpulan.

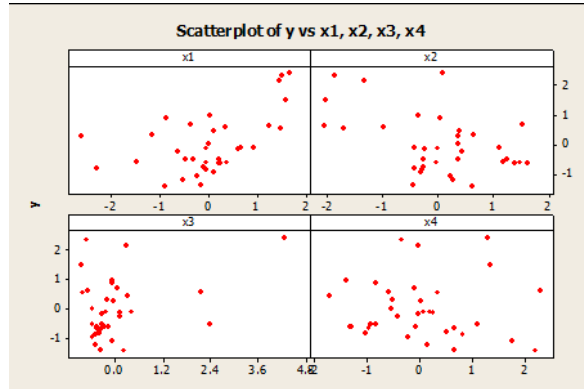
HASIL DAN PEMBAHASAN

1. Pemodelan Indeks Pembangunan Manusia Kabupaten/ Kota di provinsi Jawa Tengah dengan Pendekatan Regresi Nonparametrik Spline.

Pemodelan Indeks Pembangunan Manusia menurut kabupaten/kota di Provinsi Jawa Tengah dengan pendekatan spline dengan 4 variabel respon. Dalam membuat model Indeks Pembangunan Manusia di Provinsi Jawa Tengah langkah-langkahnya yaitu:

a. Plot data Indeks Pembangunan Manusia dan variabel prediktor

Sebelum melakukan pemodelan Indeks Pembangunan Manusia (IPM) dengan faktor-faktor yang mempengaruhinya, terdapat beberapa langkah yang perlu dilakukan antara lain melihat hubungan antara IPM dengan masing-masing variabel prediktor dan juga perlu terlebih dahulu melihat hubungan antara IPM dengan masing-masing variabel prediktor. Pada Gambar berikut menunjukkan hubungan antara IPM Jawa Tengah dengan 4 variabel prediktor yang mempengaruhinya.



Gambar Scaterplot Variabel Y dan X

Berdasarkan Gambar diatas yang menampilkan scaterplot antra variabel yang diduga berpengaruh dapatdilihat bahwa data pada scaterplot pola data menyebar, tidak menunjukkan sebuah pola tertentu dan berubah-ubah pada sub-sub interval tertentu. Oleh karena itu, dalam penelitian ini, akan digunakan regresi nonparametrik.

b. Pemodelan IPM dengan Metode GCV dan UBR

Pada pemodelan dengan metode GCV dalam pemilihan titik knot optimal nya menggunakan titik knot 1 , 2 dan 3 titik knot optimal.

1) Pemilihan Titik Knot Optimal dengan Satu Titik Knot

Berikut ini merupakan estimasi model regresi nonparametrik Spline dengan satu titik knot pada kasus IPM Jawa Tengah.

$$y = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 (x_1 - K_1)_+^1 + \hat{\gamma}_3 x_2 + \hat{\gamma}_4 (x_2 - K_2)_+^1 + \hat{\gamma}_5 x_3 + \hat{\gamma}_6 (x_3 - K_3)_+^1 + \hat{\gamma}_7 x_4 + \hat{\gamma}_8 (x_4 - K_4)_+^1$$

Berikut adalah hasil dengan menggunakan satu titik knot.

Tabel 1 Pemilihan Titik Knot Optimum Satu Titik Knot Metode GCV

GCV	X1	X2	X3	X4
0.4312*	0.5074	0.6426	2.8768	1.2206
0.4348	0.5947	0.7185	2.9808	1.3023
0.4366	0.6820	0.7943	3.0848	1.3841
0.4339	0.7693	0.8701	3.1889	1.4658
0.4357	0.8566	0.9459	3.2929	1.5476
0.4426	0.9439	1.0218	3.3969	1.6294
0.4483	1.0312	1.0976	3.5009	1.7111
0.4553	1.1185	1.1734	3.6049	1.7929
0.4698	1.2058	1.2493	3.7089	1.8746
0.4763	1.2931	1.3251	3.8130	1.9564

*) Nilai GCV Minimum

Berdasarkan table 1 nilai GCV paling minimum adalah 0.4312 dengan titik knot optimum masing-masing variabel adalah sebagai berikut: $K_1=0.5074$; $K_2=0.6426$; $K_3=2.8768$; $K_4=1.2206$ knot akan dibandingkan dengan hasil dari GCV dan UBR dengan menggunakan dua titik knot, dan tiga titik knot, perbandingan hasil GCV dan UBR tersebut dilakukan untuk memperoleh nilai GCV yang paling minimum dan diharapkan dapat menghasilkan metode *spline* terbaik.

Selanjutnya akan dicari titik knot optimal menggunakan metode UBR dimana hasilnya disajikan pada Tabel dibawah ini



Tabel 2 Pemilihan Titik Knot Optimum Satu Titik Knot Metode UBR

UBR	X1	X2	X3	X4
4.3461×10^{-12}	-2.5478	-2.0113	-0.7637	-1.6408
3.3339×10^{-13}	-2.4605	-1.9355	-0.6596	-1.5590
1.8170×10^{-13}	-2.3732	-1.8596	-0.5556	-1.4773
2.4570×10^{-16}*	1.0312	1.0976	3.5009	1.7111
1.8918×10^{-15}	1.1185	1.1734	3.6049	1.7929
5.5214×10^{-16}	0.4201	0.5668	2.7728	1.1388
7.0722×10^{-16}	0.5074	0.6426	2.8768	1.2206
7.0520×10^{-16}	0.5947	0.7185	2.9808	1.3023
4.9310×10^{-16}	0.6820	0.7943	3.0848	1.3841
7.6432×10^{-16}	0.7693	0.8701	3.1889	1.4658

*) Nilai UBR Minimum

Berdasarkan Tabel 2 didapatkan nilai UBR optimal untuk satu titik knot, yaitu 2.4570×10^{-16} * dengan titik knot optimum yang didapat $K_1= 1.0312$; $K_2 = 1.0976$; $K_3= 3.5009$; $K_4= 1.711$ knot akan dibandingkan dengan hasil dari GCV, CV dan UBR dengan menggunakan dua titik knot, dan tiga titik knot, perbandingan hasil GCV, CV dan UBR tersebut dilakukan untuk memperoleh nilai GCV, CV dan UBR yang paling minimum dan diharapkan dapat menghasilkan metode *spline* terbaik.

Selanjutnya akan dicari titik knot optimal menggunakan metode GCV dan UBR dengan titik knot 2 dan 3.

2) Pemilihan Titik Knot Optimal dengan Dua Titik Knot

Adapun model regresi nonparametrik *Spline* dengan menggunakan dua titik knot pada variabel-variabel yang mempengaruhi Indeks Pembangunan Manusia adalah sebagai berikut.

Berikut adalah hasil perhitungan GCV dan UBR untuk regresi nonparametrik *Spline* dengan menggunakan dua titik knot.

Tabel 3 Pemilihan Titik Knot Optimum Dua Titik Knot Metode GCV

GCV	X1	X2	X3	X4
0.4598	0.6917	1.1670	0.8027	1.2156
	3.0964	3.6627	1.3932	1.8383
0.4758	0.2164	1.1670	0.3899	1.2156
	2.5301	3.6627	0.9481	1.8383
0.3121*	-1.6846	-1.2093	-1.2614	-0.8486
	0.2648	0.8312	-0.8323	-0.3872
0.5125	0.2164	0.6917	0.3899	0.8027
	2.5301	3.0964	0.9481	1.3932
0.5693	-0.2588	1.1670	-0.0229	1.2156
	1.9638	3.6627	0.5029	1.8383
0.5684	-2.1598	-1.6846	-1.6743	-1.2614
	-0.3014	0.2648	-1.2774	-0.8323
0.3738	-1.6846	0.2164	-1.2614	0.3899
	0.2648	2.5301	-0.8323	0.9481
0.3297	-1.6846	1.1670	-1.2614	1.2156
	0.2648	3.6627	-0.8323	1.8383



0.3478	-1.6846	-0.7340	-1.2614	-0.4357
	0.2648	1.3975	-0.8323	0.0578

*) Nilai GCV Minimum

Berdasarkan tabel 3 nilai GCV paling minimum adalah 0.3121. Selanjutnya akan dicari titik knot optimal menggunakan metode UBR dimana hasilnya disajikan pada Tabel 4

Tabel 4 Pemilihan Titik Knot Optimum Dua Titik Knot Metode UBR

GCV	X1	X2	X3	X4
6.5180 x 10 ⁻¹⁶	-2.6351	0.8566	-2.0871	0.9459
	-0.8677	3.2929	-1.7225	1.5476
5.2885 x 10 ⁻¹⁶	-2.6351	0.9439	-2.0871	1.0218
	-0.8677	3.3969	-1.7225	1.6294
2.1471 x 10 ⁻¹⁶	-2.6351	1.0312	-2.0871	1.0976
	-0.8677	3.5009	-1.7225	1.7111
9.1801 x 10⁻¹⁷*	-2.6351	1.1185	-2.0871	1.1734
	-0.8677	3.6049	-1.7225	1.7929
1.1973 x 10 ⁻¹⁶	-2.6351	1.2058	-2.0871	1.2493
	-0.8677	3.7089	-1.7225	1.8746
3.4990 x 10 ⁻¹⁵	-2.6351	1.2931	-2.0871	1.3251
	-0.8677	3.8130	-1.7225	1.9564
3.2916 x 10 ⁻¹⁶	-2.6351	1.3803	-2.0871	1.4009
	-0.8677	3.9170	-1.7225	2.0381
1.4301 x 10 ⁻¹⁶	-2.6351	1.4676	-2.0871	1.4767
	-0.8677	4.0210	-1.7225	2.1199
6.6320 x 10 ⁻¹⁶	-2.6351	1.5549	-2.0871	1.5526
	-0.8677	4.1250	-1.7225	2.2017

*) Nilai UBR Minimum

Berdasarkan tabel 4 nilai UBR paling minimum adalah 9.1801 x 10⁻¹⁷.

3) Pemilihan Titik Knot Optimal dengan Tiga Titik Knot

Adapun model regresi nonparametrik *Spline* dengan menggunakan tiga titik knot pada variabel-variabel yang mempengaruhi Indeks Pembangunan Manusia adalah sebagai berikut.

$$\hat{y} = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 (x_1 - K_1) + \hat{\gamma}_3 (x_1 - K_2) + \hat{\gamma}_4 (x_1 - K_3) + \hat{\gamma}_5 x_2 + \hat{\gamma}_6 (x_2 - K_4) + \hat{\gamma}_7 (x_2 - K_5) + \hat{\gamma}_8 (x_2 - K_6) + \hat{\gamma}_9 x_3 + \hat{\gamma}_{10} (x_3 - K_7) + \hat{\gamma}_{11} (x_3 - K_8) + \hat{\gamma}_{12} (x_3 - K_9) + \hat{\gamma}_{13} x_4 + \hat{\gamma}_{14} (x_4 - K_{10}) + \hat{\gamma}_{15} (x_4 - K_{11}) + \hat{\gamma}_{16} (x_4 - K_{12})$$

Pada persamaan diatas dapat dilihat bahwa untuk masing-masing variabel prediktor dibutuhkan tiga titik knot. Sama halnya dengan menggunakan satu titik knot dan dua titik knot, untuk memperoleh knot yang optimum dipilih melalui nilai GCV dan UBR yang paling minimum. Berikut adalah hasil perhitungan GCV dan UBR untuk regresi nonparametrik *Spline* dengan menggunakan tiga titik knot.

Tabel 5 Pemilihan Titik Knot Optimum Tiga Titik Knot Metode GCV

GCV	X1	X2	X3	X4
0.2157	-1.9368	-1.7622	1.1185	-1.4805
	-1.3288	1.1734	-0.0355	0.1724
	3.6049	-1.0685	-0.9050	1.7929
0.2084*	-1.9368	-1.7622	1.0312	-1.4805
	-1.3288	1.0976	-0.0355	0.1724



	3.5009	-1.0685	-0.9050	1.7111
	-1.9368	-1.7622	1.2058	-1.4805
0.2418	-1.3288	1.2493	-0.0355	0.1724
	3.7089	-1.0685	-0.9050	1.8746
	-1.9368	-1.7622	1.2931	-1.480
0.2710	-1.3288	1.3251	-0.0355	0.1724
	3.8130	-1.0685	-0.9050	1.9564
	-1.9368	-1.7622	0.8566	-1.4805
0.2143	-1.3288	0.9459	-0.0355	0.1724
	3.2929	-1.0685	-0.9050	1.5476

*) Nilai GCV minimum

Berdasarkan Tabel 5 nilai GCV minimum yang diperoleh adalah 0.2084. Selanjutnya akan dicari titik knot optimal menggunakan metode UBR dimana hasilnya disajikan pada Tabel 6

Tabel 6 Pemilihan Titik Knot Optimum Tiga Titik Knot Metode UBR

CV	X1	X2	X3	X4
	-1.5003	1.11851	1.3803	-1.1013
1.0782 x 10 ⁻¹⁴	1.1734	1.4009	0.4844	3.6049
	3.9170	-0.6597	1.7929	2.0381
	-1.5003	1.11851	1.4676	-1.1013
1.0421 x 10 ⁻¹⁴	1.1734	1.4767	0.4844	3.6049
	4.0210	-0.6597	1.7929	2.1199
	-1.5003	1.1185	1.5549	-1.1013
8.714 x 10^{-15*}	1.1734	1.5526	0.4844	3.6049
	4.1250	-0.6597	1.7929	2.2017
	-1.5003	1.2058	1.2931	-1.1013
1.9448 x 10 ⁻¹⁴	1.2493	1.3251	0.4844	3.7089
	3.8130	-0.6597	1.8746	1.9564
	-1.5003	1.2058	1.3803	-1.1013
4.8336 x 10 ⁻¹⁴	1.2493	1.4009	0.4844	3.7089
	3.9170	-0.6597	1.8746	2.0381

*) Nilai UBR minimum

Berdasarkan Tabel 6 nilai UBR minimum yang diperoleh adalah 8.714 x 10^{-15*}

c. Perbandingan Nilai Titik Knot GCV, CV dan UBR

Setelah mendapatkan nilai GCV dan UBR dari pemodelan satu knot, dua knot dan tiga knot selanjutnya adalah membandingkan nilai GCV dan UBR untuk memilih knot mana yang terbaik. Berikut ini adalah nilai GCV dan UBR dari masing –masing knot.

Tabel 7 Perbandingan Nilai GCV, CV dan UBR Minimum

Titik knot	GCV	UBR
Knot 1	0.4312	2.4570 x 10 ⁻¹⁶
Knot 2	0.3121	9.1801 x 10⁻¹⁷
Knot 3	0.2084	8.714 x 10 ⁻¹⁵

Pada tabel 7 dapat dilihat bahwa pada metode GCV titik knot yang terkecil terdapat pada titik knot 3 yang menghasilkan nilai minimum sebesar 0.2084, sedangkan dengan



metode UBR titik knot terkecil terdapat pada titik knot 2 menghasilkan nilai minimum sebesar 9.1801×10^{-17}

d. Estimasi Parameter

Nilai estimasi parameter metode GCV untuk *Spline Truncated* 3 knot yaitu sebagai berikut:

Tabel 8 Nilai Estimasi Parameter GCV

Parameter	Estimasi	Parameter	Estimasi
β_0	-1.9612	β_9	2.9222
β_1	3.7031	β_{10}	-8.5965
β_2	-19.1771	β_{11}	6.0100
β_3	15.7586	β_{12}	-1.9890
β_4	5.9038	β_{13}	-3.7115
β_5	-3.6425	β_{14}	16.5261
β_6	12.0020	β_{15}	-13.6162
β_7	-8.2976	β_{16}	2.3242
β_8	-0.1171		

Pemodelan yang didapatkan setelah melihat dari nilai GCV paling minimum yaitu model *spline truncated* dengan 3 knot sehingga didapatkan model sebagai berikut :

$$\hat{y} = -1.9612 + 3.7031x_1 - 19.1771(x_1 - 1.9368) + 15.7586(x_1 - 1.3288) + 5.9038(x_1 - 3.5009) - 3.6425x_2 + 12.0020(x_2 - 1.7622) - 8.2976(x_2 - 1.0976) - 0.1171(x_2 - 1.0685) + 2.9222x_3 - 8.59659(x_3 - 1.0312) + 6.0100(x_3 - 0.0355) - 1.9890(x_3 - 0.9050) - 3.71159x_4 + 16.5261(x_4 - 1.4805) - 13.6162(x_4 - 0.1724) + 2.3242(x_4 - 1.711)$$

Berdasarkan analisis yang telah digunakan sebelumnya, maka didapatkan model regresi nonparametrik *spline truncated* yang terbaik adalah dengan menggunakan 3 knot. Nilai koefisien determinasi yang dihasilkan oleh model regresi nonparametrik *spline truncated* adalah sebesar 94.32 persen. Nilai tersebut menunjukkan bahwa variabel Indeks Pembangunan Manusia di Provinsi Jawa Tengah pada tahun 2017 dapat dijelaskan oleh keempat variabel prediktor adalah sebesar 94.32 persen sedangkan sisanya 5,68 persen dijelaskan oleh variabel lain yang tidak terdapat pada model.

Selanjutnya nilai estimasi parameter metode UBR untuk *spline truncated* dengan 2 titik knot sebagai berikut:

Tabel 9 Nilai Estimasi Parameter UBR

Parameter	Estimasi	Parameter	Estimasi
β_0	1.2415	β_6	-0.6925
β_1	-0.6394	β_7	0.8073
β_2	1.3023	β_8	2.3971
β_3	3.3901	β_9	-0.7533
β_4	-0.6925	β_{10}	1.8314



$$\begin{array}{cc} \beta_5 & -1.2533 & \beta_{11} & -3.0994 \\ & & \beta_{12} & 1.1103 \end{array}$$

Pemodelan yang didapatkan setelah melihat dari nilai UBR paling minimum yaitu model *Spline Truncated* dengan 2 knot sehingga didapatlah model sebagai berikut :

$$y = 1.2415 - 0.6394x_1 + 1.3023(x_1 - 2.6351) + 3.3901(x_1 - 0.8677) - 0.6925x_2 - 1.2533(x_2 - 1.1185) - 0.6925(x_2 - 3.6049) + 0.8073x_3 + 2.3971(x_3 - 2.0871) - 0.7533(x_3 - 1.7225) + 1.8314x_4 - 3.0994(x_4 - 1.1734) + 1.1103(x_4 - 1.7929)$$

Berdasarkan analisis yang telah digunakan sebelumnya, maka didapatkan model regresi nonparametrik *spline truncated* yang terbaik adalah dengan menggunakan 2 knot. Nilai koefisien determinasi yang dihasilkan oleh model regresi nonparametrik *spline truncated* adalah sebesar 74.12 persen. Nilai tersebut menunjukkan bahwa variabel Indeks Pembangunan Manusia di Provinsi Jawa Tengah pada tahun 2017 dapat dijelaskan oleh keempat variabel prediktor adalah sebesar 74.13 persen sedangkan sisanya 25.87 persen dijelaskan oleh variabel lain yang tidak terdapat pada model.

e. Pemilihan Model Terbaik

Untuk melihat metode mana yang menghasilkan model yang paling baik, dapat dilihat dari nilai Mean Square Error (MSE) dan R^2 sebagai kriteria kebaikan model. Nilai MSE dan R^2 untuk pemilihan titik knot optimal GCV dan UBR disajikan pada tabel 17.

Tabel 10 Perbandingan Nilai MSE dan R^2 Metode GCV dan UBR

	GCV	UBR
MSE	0.107	0.399
R^2	94.324	74.132

Berdasarkan table 10 terlihat bahwa metode GCV menghasilkan nilai MSE yang lebih kecil dan nilai R^2 yang lebih besar dibandingkan dengan metode UBR yang mana nilai R^2 sebesar 94.324% sedangkan nilai R^2 UBR sebesar 74.132%.

Hal ini berarti metode GCV lebih baik digunakan dibandingkan dengan metode UBR pada pemodelan Indeks Pembangunan Manusia di Jawa Tengah pada tahun 2017.

f. Interpretasi Model

Interpretasi model terdapat variabel-variabel yang berpengaruh secara signifikan sebagai berikut:

1. Hubungan antara Angka Melek Huruf (X1) terhadap Indeks Pembangunan Manusia (Y) sebagai berikut:

$$\hat{y} = 3.7031x_1 - 19.1771(x_1 - 1.9368) + 15.7586(x_1 - 1.3288) + 5.9038(x_1 - 3.5009)$$

$$\left\{ \begin{array}{ll} 3.7071x_1 & ; x_1 < 1.9368 \\ -15.4740x_1 + 37.1422 & ; 1.9368 \leq x_1 < 1.3288 \\ 0.2846x_1 - 20.9400 & ; 1.3288 \leq x_1 < 3.5009 \\ 6.1884x_1 - 20.6686 & ; x_1 \geq 3.5009 \end{array} \right.$$

2. Hubungan antara Jumlah Fasilitas Kesehatan (X2) terhadap Indeks Pembangunan Manusia (Y) sebagai berikut:

$$\hat{y} = -3.6425x_2 + 12.0020(x_2 - 1.7622) - 8.2970(x_2 - 1.0976) - 0.1171(x_2 - 1.0685)$$



$$\begin{cases} -3.6425x_2 & ; x_2 < 1.7622 \\ 8.3595x_2 - 21.1499 & ; 1.7622 \leq x_2 < 1.0976 \\ 0.0625x_2 + 9.1067 & ; 1.0976 \leq x_2 < 1.0685 \\ -0.0546x_2 + 0.1251 & ; x_2 \geq 1.0685 \end{cases}$$

3. Hubungan antara PDRB (X3) terhadap Indeks Pembangunan Manusia (Y) sebagai berikut:

$$\hat{y} = 2.9222x_3 - 8.5965(x_3 - 1.0312) + 6.0100(x_3 - 0.0355) - 1.9890(x_3 - 0.9050)$$

$$\begin{cases} 2.9222x_3 & ; x_3 < 1.0312 \\ -5.6740x_3 + 8.8647 & ; 1.0312 \leq x_3 < 0.0355 \\ 0.3360x_3 - 0.2133 & ; 0.0355 \leq x_3 < 0.9050 \\ -1.6530x_3 + 1.8000 & ; x_3 \geq 0.9050 \end{cases}$$

4. antara Tingkat Pengangguran Terbuka (TPT) terhadap Indeks Pembangunan Manusia (Y) sebagai berikut:

$$\hat{y} = -3.7115x_4 + 16.5261(x_4 - 1.4805) - 13.6162(x_4 - 0.1724) + 2.3242(x_4 - 1.711)$$

$$\begin{cases} -3.7115x_4 & ; x_4 < 1.4805 \\ -20.2376x_4 - 24.4668 & ; 1.4805 \leq x_4 < 0.1724 \\ -33.8538x_4 + 2.3474 & ; 0.1724 \leq x_4 < 2.3243 \\ -31.5296x_4 - 3.9767 & ; x_4 \geq 2.3243 \end{cases}$$

KESIMPULAN

1. Untuk metode UBR menggunakan regresi nonparametrik *spline* terbaik yang dihasilkan adalah *spline* dengan 2 knot. Sehingga didapatkan model sebagai berikut:

$$\begin{aligned} y = & 1.2415 - 0.6394x_1 + 1.3023(x_1 - 2.6351) + 3.3901(x_1 - 0.8677) - 0.6925x_2 \\ & - 1.2533(x_2 - 1.1185) - 0.6925(x_2 - 3.6049) + 0.8073x_3 + 2.3971(x_3 - 2.0871) \\ & - 0.7533(x_3 - 1.7225) + 1.8314x_4 - 3.0994(x_4 - 1.1734) + 1.1103(x_4 - 1.7929) \end{aligned}$$

- Untuk metode GCV menggunakan regresi nonparametrik *spline* terbaik yang dihasilkan adalah *spline* dengan 3 knot. Sehingga didapatkan model sebagai berikut:

$$\begin{aligned} \hat{y} = & -1.9612 + 3.7031x_1 - 19.1771(x_1 - 1.9368) + 15.7586(x_1 - 1.3288) \\ & + 5.9038(x_1 - 3.5009) - 3.6425x_2 + 12.0020(x_2 - 1.7622) \\ & - 8.2976(x_2 - 1.0976) - 0.1171(x_2 - 1.0685) + 2.9222x_3 \\ & - 8.59659(x_3 - 1.0312) + 6.0100(x_3 - 0.0355) \\ & - 1.9890(x_3 - 0.9050) - 3.71159x_4 + 16.5261(x_4 - 1.4805) \\ & - 13.6162(x_4 - 0.1724) + 2.3242(x_4 - 1.711) \end{aligned}$$

2. Untuk pemilihan titik knot optimal pada model regresi nonparametrik *Spline Truncated* menggunakan metode GCV dan UBR pada aplikasi data Indeks Pembangunan (IPM) di Provinsi Jawa Tengah tahun 2017, model terbaik untuk metode GCV menghasilkan nilai R^2 sebesar 94.3245% dan nilai MSE sebesar 0.1072 untuk metode UBR didapatkan nilai R^2 sebesar 74.1322 sedangkan MSE sebesar 0.3997 setelah membandingkan nilai R^2 dan MSE dari ketiga metode didapatkan hasil bahwa metode GCV yang lebih baik dibandingkan metode UBR, model terbaik dengan metode GCV yang menggunakan 3 titik knot.

DAFTAR PUSTAKA



- Awal, C. P. 2014. Pendekatan Regresi Nonparametrik Spline untuk Memodelkan Indeks Pembangunan Manusia (IPM) di Papua. Surabaya: Institut Teknologi Sepuluh Nopember.
- Badan Pusat Statistik. 2018. *Indeks Pembangunan Manusia 2015-2018*. Jakarta:Badan Pusat Statistik Indonesia.
- Badan Pusat Statistik. 2018. *Jawa Tengah Dalam Angka tahun 2018*. Jakarta:Badan Pusat Statistik Indonesia.
- Badan Pusat Statistik, 2017. *Statistik Pendidikan Provinsi Jawa Tengah 2017*.BPS Jawa Tengah, Semarang.
- Badan Pusat Statistik, 2017. *Laju Pertumbuhan PDRB Atas Harga Berlaku Konstan Menurut Kabupaten/Kota*. Provinsi Jawa Tengah 2017.BPS Jawa Tengah, Semarang.
- Badan Pusat Statistik, 2017. *Tingkat Pengangguran Terbuka Menurut Kabupaten/kota Provinsi Jawa Tengah 2017*.BPS Jawa Tengah, Semarang.
- Budiantara, I. N 2005. *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametric*. Berkala MIPA, Universitas Gadjah Mada, Yogyakarta.
- Budiantara, I. N. 2006, Regresi Nonparametrik Dalam Statistika, Makalah Pembicara Utama pada Seminar Nasional Matematika. Makassar : Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Makasar (UNM).
- Budiantara, I. N 2009. *Spline dalam Regresi Nonparametrik dan Semiparametrik : sebuah pemodelan Masa Kini dan Masa Mendatang*. Surabaya: ITS Press.
- Devi, Alvita Rachmana, 2015, Metode Unbiased Risk dan Cross Validation untuk pemilihan Titik Knot Optimal dalam Regresi Nonparametrik Multivariabel Spline Truncated, Tesis, Statistika, Institut Teknologi Sepuluh November, Surabaya.
- Draper, N.R. and Smith, H. 1992. *Applied Regression Analysis, Second Edition*. New York : John Wiley and sons, Inc.
- Eubank, R.L. 1999. *Nonparametric Regression and Spline Smoothing Second Edition*. New York : Marcel Dekker.
- Fox, J., 2008, *Applied Regression Analysis and Generalized Linier Models*, Sage Publications, London.
- Melliana, 2013. Analisis Statistika Faktor yang Mempengaruhi IPM di kabupaten/kota provinsi Jawa Timur dengan Menggunakan regresi panel, Institut Teknologi Sepuluh Nopember, Surabaya.
- Sari, Sulistya Umie Ruhmana, 2016, Perbandingan odel Spline Multivariabel dengan menggunakan metode *Generalized Cross Validation* dan *Unbiasses Risk* dalam Pemilihan Titik Knot Optimal. Tesis, Statistika, Institut Teknologi Sepuluh November, Surabaya.
- S. Asmin, 2010, *Pemodelan Nilai Unas IPA dengan Pendekatan Regresi Semiparametrik Spline di SMAN 1 Grati Pasuruan*, Institut Teknologi Sepuluh Nopember Surabaya.
- Wahba.G. 1990. *Spline Models For Observational Data*. University Of Winsconsin at Madison.
- Wang, Y., 1998, "Spline Smoothing Models with Correlated Error", *Journal of the Royal Statistical Sociaty, Series B*, 5r0, 341-348.
- Wang, Y., 2011, *Smoothing Splines Methods and Applications*, CRC Press, California