

## PENGGUNAAN METODE *MEAN SQUARE ERROR* PADA REGRESI *SPLINE* ORDE 2

Agustini Tripena\*

Jurusan Matematika, Universitas Jenderal Soedirman  
agustini.brsurbakti@unsoed.ac.id

**ABSTRACT.** *The mean square error is the estimated value of the residual variance. The model with the minimum MSE is the best model, meaning that the estimated value is close to the actual value. Indeks pembangunan manusia (IPM) is an illustration to see the level of human welfare. In Central Java Province, the IPM figure has reached 72.16 percent in 2020, so indeks pembangunan manusia is a big influence for the government. Many factors affect the IPM in Central Java and it is not known which factors influence the most. Spline regression is one method that can be used, with its segmented nature allowing it to adapt effectively to the characteristics of the data. Quadratic spline regression was used to estimate Central Java IPM on 4 influencing factors with 35 observations each and minimum MSE criteria. This study shows that the percentage of life expectancy at birth ( $X_1$ ), expected years of schooling ( $X_2$ ), average length of schooling ( $X_3$ ) and per capita expenditure ( $X_4$ ) are the factors that most influence Central Java's IPM. The number of knots used is 12 knots with  $K_1 = 76,6$ ;  $K_2 = 75,78$ ;  $K_3 = 73,40$  is on the variable  $X_1$ ,  $K_4 = 13,5$ ;  $K_5 = 13,05$ ;  $K_6 = 13,31$  is on the variable  $X_2$ ,  $K_7 = 10,39$ ;  $K_8 = 8,56$ ;  $K_9 = 6,97$  is on the variable  $X_3$ ,  $K_{10} = 12210$ ;  $K_{11} = 9263$ ;  $K_{12} = 12467$  is on the variable  $X_4$ . The resulting MSE is 0,01181046 with a determinant coefficient of 98,8861%.*

**Keywords:** *Mean Square Error, quadratic spline regression, knot point, IPM*

**ABSTRAK.** *Mean square error* merupakan nilai taksiran dari variansi residua. Model dengan MSE minimum menjadi model terbaik, artinya nilai taksiran mendekati nilai sebenarnya. Indeks pembangunan manusia (IPM) menjadi gambaran untuk melihat tingkat kesejahteraan manusia. Di Provinsi Jawa Tengah angka IPM kurang lebih sudah mencapai 72,16 persen pada tahun 2020, sehingga indeks pembangunan manusia menjadi pengaruh besar bagi pemerintah. Banyak faktor yang mempengaruhi IPM Jawa Tengah dan tidak diketahui faktor mana yang paling mempengaruhi. Regresi *spline* adalah salah satu metode yang dapat digunakan, dengan sifat tersegmen yang dimilikinya memungkinkan menyesuaikan diri secara efektif terhadap karakteristik data. Regresi *spline* kuadratik digunakan untuk mengestimasi IPM Jawa Tengah terhadap 4 faktor yang mempengaruhi dengan masing-masing sebanyak 35 data amatan dan kriteria MSE minimum. Penelitian ini menunjukkan bahwa presentase umur harapan hidup saat lahir ( $X_1$ ), harapan lama sekolah ( $X_2$ ), rata-rata lama sekolah ( $X_3$ ) dan pengeluaran per kapita ( $X_4$ ) merupakan faktor yang paling mempengaruhi IPM Jawa Tengah. Jumlah titik knot yang digunakan sebanyak 12 titik knot dengan  $K_1 = 76,6$ ;  $K_2 = 75,78$ ;  $K_3 = 73,40$  berada pada variabel  $X_1$ ,  $K_4 = 13,5$ ;  $K_5 = 13,05$ ;  $K_6 = 13,31$  berada pada variabel  $X_2$ , ;  $K_7 = 10,39$ ;  $K_8 = 8,56$ ;  $K_9 = 6,97$  berada pada variabel  $X_3$ ,  $K_{10} = 12210$ ;  $K_{11} = 9263$ ;  $K_{12} = 12467$

berada pada variabel  $X_4$ , MSE yang dihasilkan sebesar 0,01181046 dengan koefisien determinan sebesar 98,8861%.

**Kata kunci:** *Mean Square Error*, regresi *spline* kuadrat, titik knot

## 1. PENDAHULUAN

Pulau Jawa merupakan pulau dengan jumlah penduduk yang paling padat, salah satu provinsi yang paling padat penduduknya adalah Jawa Tengah, dengan semakin padat penduduk dalam satu wilayah maka angka Indeks Pembangunan Manusia (IPM) harus diperhitungkan. 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.

IPM menjadi indikator penting untuk mengukur salah satu aspek penting yang berkaitan dengan kualitas dari hasil pembangunan ekonomi, yakni derajat perkembangan manusia. Dalam kata lain, terdapat suatu korelasi positif antara nilai IPM dengan derajat keberhasilan pembangunan ekonomi (Tambunan, 2003 dalam Nurmalisa dan Nur, 2019). Oleh karena itu perlu diketahui faktor-faktor apa saja yang mempengaruhi IPM di Provinsi Jawa tengah. Salah satu metode alternatif lain untuk memodelkan serta menganalisis IPM yang digunakan untuk mengetahui faktor-faktor yang berpengaruh terhadap IPM di Jawa Tengah adalah dengan analisis regresi *spline*.

Regresi *spline* merupakan salah satu metode dalam regresi nonparametrik yang digunakan untuk mengestimasi bentuk kurva yang tidak terikat dengan asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang baik dalam mengestimasi kurva (Eubank, R. 1988). Dalam pendekatan kurva, regresi *spline* tetap memperhatikan kemulusan kurva dimana kurva termuat dalam suatu fungsi tertentu. *Spline* merupakan model polinomial yang tersegmen, sifat tersegmen inilah yang memberikan fleksibilitas yang lebih baik dari pada model polinomial biasa (Eubank, 1988). Sifat ini memungkinkan model regresi *spline* menyesuaikan diri secara efektif terhadap karakteristik lokal dari data.

Regresi *spline* linier (orde 1) biasanya diaplikasikan pada data dengan pola yang masih sederhana sedangkan regresi *spline* kuadratik (orde 2) dan regresi *spline* kubik (orde 3) biasanya diaplikasikan pada pola data yang lebih kompleks. Menurut Ruppert et al. (2003) regresi *spline* kuadratik memiliki kelebihan untuk menyesuaikan diri terhadap titik ekstrem pada plot data khususnya pada data yang memiliki pola kuadrat. Regresi *spline* kuadratik dapat diterapkan dalam berbagai bidang, baik dalam bidang kesehatan, ekonomi, demografi, dan lainnya. Hal ini dapat dilihat pada penelitian yang dilakukan Beniar B (2016) yang memodelkan *gross domestic product* dan Tripena (2011) yang memodelkan perubahan konduktasi. Menurut penulis, metode regresi *spline* kuadratik tepat digunakan untuk memodelkan IPM di Jawa Tengah karena data cenderung bervariasi dan membentuk pola kuadrat.

*Mean square error* (MSE) merupakan nilai taksiran dari variansi residual. Model dengan MSE yang minimum menjadi model terbaik, artinya nilai taksiran mendekati nilai sebenarnya. Pemilihan lokasi titik-titik knot sangatlah penting untuk mendapat model regresi *spline* terbaik. Salah satu metode yang sering digunakan untuk pemilihan lokasi titik knot untuk memilih model regresi terbaik dapat menggunakan nilai *mean square error* (MSE). Menurut Eubank (1988), untuk menyelesaikan pemilihan parameter dengan meminimalkan MSE atau metode lainnya pada variasi yang sangat luas. Semakin kecil nilai MSE pada titik knot maka estimasi model regresi *spline* akan semakin baik. Perhitungan MSE ini juga akan digunakan dalam menentukan banyaknya titik knot yang akan digunakan.

Mengacu pada riset diatas, penulis tertarik untuk mengkaji bagaimana penggunaan MSE pada regresi *spline* orde 2 dengan titik knot pada data IPM Provinsi Jawa Tengah pada tahun 2020 (Badan Pusat Statistika, 2020).

## 2. METODE PENELITIAN

Pada penelitian ini, metode yang akan digunakan adalah studi kasus dengan menggunakan data sekunder tentang IPM di Jawa Tengah tahun 2020

yang diperoleh dari BPS Jawa Tengah. Data yang diperoleh dalam penelitian ini akan dianalisis regresi *spline* orde dua dengan bantuan *software R* dan *Minitab 16*.

Tahapan-tahapan yang dilakukan dalam analisis data adalah sebagai berikut:

- a. Membuat scatter plot antara variabel prediktor dengan variabel respon untuk mengetahui pola data.
- b. Memilih lokasi serta banyaknya titik knot dan faktor-faktor yang mempengaruhi sekaligus penempatan titik-titik knot tersebut,
- c. Membentuk model regresi *spline* orde dua untuk IPM dan faktor-faktor yang mempengaruhi dengan menggunakan beberapa titik knot sekaligus penempatan titik-titik knot tersebut, yakni: Model regresi *spline* orde 2 (kuadrat) dengan tiga titik knot:

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 (x_i - K_1)_+^2 + \beta_4 (x_i - K_2)_+^2 + \beta_5 (x_i - K_3)_+^2$$

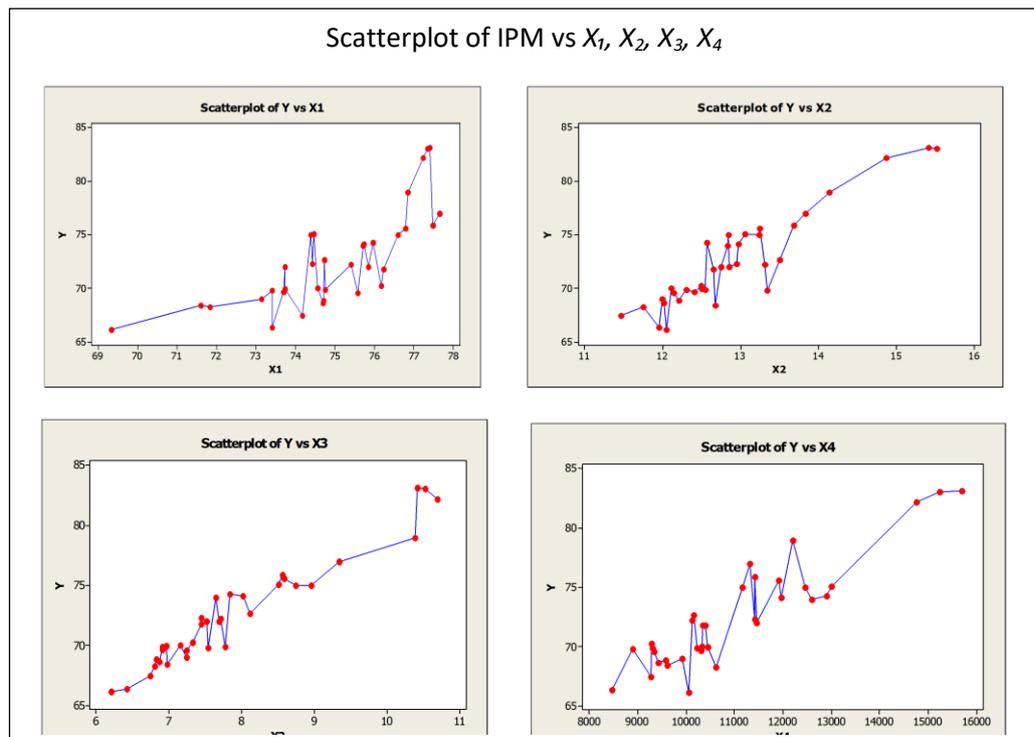
- d. Menghitung nilai  $MSE(\lambda)$  untuk masing-masing model regresi *spline*. serta memilih model regresi *spline* terbaik diantara model-model yang didapatkan berdasarkan kriteria  $MSE(\lambda)$  minimum.
- e. Menghitung nilai koefisien determinasi ( $R^2$ ) dan melakukan pengujian signifikansi parameter dan pengujian asumsi residual terbaik pada regresi *spline*.
- f. Melakukan interpretasi model dan menarik kesimpulan.
- g. Variabel-variabel yang digunakan pada penelitian:
  - $Y$  : Pencapaian indeks pembangunan Provinsi Jawa Tengah
  - $X_1$  : Umur harapan hidup saat lahir (dalam satuan tahun)
  - $X_2$  : Harapan lama sekolah (dalam satuan tahun)
  - $X_3$  : Rata-rata lama sekolah (dalam satuan tahun)
  - $X_4$  : Pengeluaran per kapita (ribu rupiah/orang/tahun)

### 3. HASIL DAN PEMBAHASAN

Membuat *scatterplot*, kemudian memilih titik knot dan estimasi model secara parsial terdapat  $(n - 2)$  titik knot yang dapat digunakan untuk membentuk model regresi *spline* tanpa mengikutkan nilai maksimum dan minimum pada data. Masing-masing variabel prediktor memiliki banyak data sebanyak 35 sehingga terdapat 33 data yang dapat digunakan sebagai titik knot. Pada penelitian ini jumlah titik knot maksimum yang digunakan sebanyak 5 titik knot, kombinasi titik knot yang digunakan yaitu titik knot yang menghasilkan nilai MSE terkecil. Variabel prediktor diasumsikan saling bebas sehingga pemilihan titik knot dapat dilakukan secara parsial.

#### 3.1 Pembuatan *Scatterplot* Seluruh Variabel Prediktor dengan Variabel Respon untuk IPM vs $X_1, X_2, X_3, X_4$

Perubahan pola data yang naik dan turun secara halus dapat menggunakan pendekatan orde kuadrat. Berikut adalah *scatterplot* dari data indeks pembangunan manusia (Y) terhadap  $X_1, X_2, X_3, X_4$ , menunjukkan perubahan pola naik dan turun secara halus pada interval tertentu.



Gambar 1. *Scatterplot* IPM dengan  $X_1, X_2, X_3, X_4$

### 3.2 Pemilihan Titik Knot dan Estimasi Model Secara Parsial

Lokasi titik knot serta banyaknya titik knot yang digunakan sangat mempengaruhi model regresi *spline* ordo dua yang dibentuk. Lokasi titik knot yang berbeda akan menghasilkan model regresi *spline* yang berbeda pula. Penentuan lokasi titik knot juga akan menentukan besar kesalahan model regresi *spline* ordo dua yang dibentuk. Memilih lokasi titik knot serta banyak jumlah titik knot optimum yang menghasilkan nilai MSE paling minimum. Berikut perolehan nilai MSE serta pemilihan lokasi titik knot dan jumlah knot yang digunakan antara 3 dan 5 knot. Pemilihan titik knot secara parsial disajikan pada langkah berikut.

**Tabel 1.** Lokasi titik knot dan nilai *MSE* variabel  $X_1, X_2, X_3, X_4$

No	Jumlah	Knot					GCV
		K1	K2	K3	K4	K5	
1		76,6	76,16	73,72	-	-	5,640152
2		76,6	76,22	77,47	-	-	5,565592
3	3	<b>76,6</b>	<b>75,78</b>	<b>73,40</b>	-	-	<b>5,054300</b>
4		76,6	76,85	73,72	-	-	5,126295
5		76,6	76,95	73,69	-	-	5,972052
6		76,22	76,85	77,4	74,41	75,71	5,551069
7		76,22	76,85	77,4	74,43	75,73	5,550695
8	5	76,22	76,85	77,4	74,47	75,93	5,566142
9		<b>76,22</b>	<b>76,85</b>	<b>77,4</b>	<b>77,18</b>	<b>75,59</b>	<b>4,659559</b>
10		76,22	76,85	77,4	74,38	77,34	5,551707
1		13,5	13,34	13,31	-	-	3,350142
2		13,5	13,83	13,31	-	-	3,346752
3	3	13,5	13,23	13,31	-	-	3,429041
4		<b>13,5</b>	<b>13,05</b>	<b>13,31</b>	-	-	<b>3,317915</b>
5		13,5	13,68	13,31	-	-	3,535496
6		13,25	13,31	13,34	12,67	13,83	2,194746
7		13,25	13,31	13,34	12,85	14,87	2,223902
8	5	<b>13,25</b>	<b>13,31</b>	<b>13,34</b>	<b>12,54</b>	<b>12,45</b>	<b>2,180,639</b>
9		13,25	13,31	13,34	14,87	13,68	2,189705
10		13,25	13,31	13,34	13,83	14,14	2,272676
1		10,39	10,42	10,53	-	-	1,457117
2		10,39	10,53	7,52	-	-	1,476217
3	3	<b>10,39</b>	<b>8,56</b>	<b>6,97</b>	-	-	<b>1,416,521</b>
4		10,39	9,34	7,24	-	-	1,492749
5		10,39	8,58	8,12	-	-	1,490814
6		<b>6,42</b>	<b>10,39</b>	<b>10,42</b>	<b>6,52</b>	<b>6,42</b>	<b>1,259531</b>

7		6,42	10,39	10,42	8.96	8,02	1,541845
8	5	6,42	10,39	10,42	9.34	10,87	1,543799
9		6,42	10,39	10,42	10.98	7,45	1,599985
10		6,42	10,39	10,42	7.52	7,44	1,596722
1		12210	12467	12467	-	-	3,659921
2		12210	9301	12467	-	-	4,271574
3	3	12210	9263	12467	-	-	3,445349
4		12210	11448	12467	-	-	4,242327
5		12210	10058	12467	-	-	4,179122
6		12999	9286	10221	11425	12910	4,099946
7		12999	9286	9266	9246	11589	3,380326
8	5	12999	9286	10328	11448	12966	4,098248
9		12999	9286	10390	11921	12210	4,003098
10		12999	9286	10343	11325	12999	4,009789

Berdasarkan Tabel 1, untuk variabel  $X_1$ , dengan 3 titik knot ketika nilai  $K_1 = 76,6$ ,  $K_2 = 75,78$ , dan  $K_3 = 73,40$  dimasukkan ke dalam program akan menghasilkan nilai MSE sebesar 5,640152, begitu pula dengan kombinasi 5 titik knot  $K_1 = 76,22$ ,  $K_2 = 76,85$ ,  $K_3 = 77,4$ ,  $K_4 = 77,18$ , dan  $K_5 = 75,59$  dengan nilai MSE minimum sebesar 4,659559. Untuk variabel  $X_2$ , dengan 3 titik knot ketika nilai  $K_1 = 13,5$ ,  $K_2 = 13,05$ , dan  $K_3 = 13,31$  dimasukkan ke dalam program akan menghasilkan nilai MSE sebesar 3,317915, begitu pula dengan kombinasi 5 titik knot  $K_1 = 13,25$ ,  $K_2 = 13,31$ ,  $K_3 = 13,34$ ,  $K_4 = 12,54$ , dan  $K_5 = 12,45$  dengan nilai MSE minimum sebesar 2,180,639. Untuk variabel  $X_3$ , dengan 3 titik knot ketika nilai  $K_1 = 10,39$ ,  $K_2 = 8,56$ , dan  $K_3 = 6,97$  dimasukkan ke dalam program akan menghasilkan nilai MSE sebesar 2,180639, begitu pula dengan kombinasi 5 titik knot  $K_1 = 6,42$ ,  $K_2 = 10,19$ ,  $K_3 = 10,42$ ,  $K_4 = 6,52$ , dan  $K_5 = 6,42$  dengan nilai MSE minimum sebesar 1,259531. Untuk variabel  $X_4$ , dengan 3 titik knot ketika nilai  $K_1 = 12210$ ,  $K_2 = 9263$ , dan  $K_3 = 12467$  dimasukkan ke dalam program akan menghasilkan nilai MSE sebesar 3,445349, begitu pula dengan kombinasi 5 titik knot  $K_1 = 12999$ ,  $K_2 = 9286$ ,  $K_3 = 9266$ ,  $K_4 = 9246$ , dan  $K_5 = 11589$  dengan nilai MSE minimum sebesar 3,380326

**Tabel 2.** Koefisien parameter model regresi spline ordo dua tiga titik knot  $X_1, X_2, X_3, X_4$

Parameter	$X_1$	$X_2$	$X_3$	$X_4$
$\beta_0$	205,1145008	236,880186	16,7180015	-211,5870
$\beta_1$	-4,0310337	-38,880438	10,3086680	0,07165321
$\beta_2$	0,0303780	1,112513	-0,3264164	-0,000003257513
$\beta_3$	5,7506581	-605,430042	2501,1763070	-0,00000001747584
$\beta_4$	-507,8600806	1552,622065	-6201,2135730	0,000001140067
$\beta_5$	1320,0285553	-1852,146613	3231,1787562	0,000001128648

Berikut diperoleh estimasi model regresi *spline* kuadratik dengan 3 titik

knot untuk variabel  $X_1, X_2, X_3, X_4$

$$\hat{Y}(X_1) = 205,1145008 - 4,0310337X_1 + 0,0303780X_1^2 + 5,7506581(X_1 - 76,6)_+^2 - 507,8600806(X_1 - 75,78)_+^2 + 1320,0285553(X_1 - 73,40)_+^2$$

$$\hat{Y}(X_2) = 236,880186 - 38,880438X_2 + 1,112513X_2^2 - 605,430042(X_2 - 13,5)_+^2 + 1552,622065(X_2 - 13,05)_+^2 - 1852,146613(X_2 - 13,31)_+^2$$

$$\hat{Y}(X_3) = 16,7180015 + 10,3086680X_3 - 0,3264164X_3^2 + 2501,1763070(X_3 - 10,39)_+^2 - 6201,2135730(X_3 - 8,56)_+^2 + 4342,1898673(X_3 - 6,97)_+^2$$

$$\hat{Y}(X_4) = -211,5870 + 0,07165321X_4 - 0,000003257513X_4^2 - 0,00000001747584(X_4 - 12210)_+^2 + 0,000001140067(X_4 - 9263)_+^2 + 0,000001128648(X_4 - 12467)_+^2$$

### 3.3 Pemilihan Titik Knot dan Estimasi Model Secara Serentak

Pemilihan titik knot dan estimasi model secara serentak berdasarkan kriteria nilai MSE yang minimum dari masing-masing variabel.

**Tabel 3.** Titik knot optimum dan nilai MSE minimum variabel prediktor

Variabel	Jumlah Knot	Knot	MSE	Koefisien Determinasi
X1	3	76,6;75,78;73,40	5,054300	0,8278634
X2	3	13,5;13,05;13,31	3,317915	0,9163762
X3	3	10,39;8,56;6,97	3,445349	0,8243003
X <sub>4</sub>	3	12210;9263;12467	1,416521	0,9460883

**Tabel 4.** Kombinasi 3 titik knot dengan 4 variabel prediktor dan nilai *MSE*

Kombinasi Variabel	Kombinasi Titik Knot	<i>MSE</i>
X <sub>1</sub> , X <sub>2</sub> , X <sub>3</sub> , X <sub>4</sub>	76,6;75,78;73,40; 13,5;13,05;13,31; 10,39;8,56;6,97; 12210;9263;12467	0,01181046

Dari Tabel 4, kriteria nilai *MSE* minimum 0,01181046 diperoleh dari kombinasi titik knot terdapat pada kombinasi variabel X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub> dan X<sub>4</sub>. Maka dapat disimpulkan bahwa model regresi *spline* ordo dua terbaik yaitu model regresi *spline* dengan menggunakan keempat variabel prediktor dengan 12 kombinasi knotnya K<sub>1</sub> = 76, 6; K<sub>2</sub> = 75,78; K<sub>3</sub> = 73,40; K<sub>4</sub> = 13,5; K<sub>5</sub> = 13,05; K<sub>6</sub> = 13,31; K<sub>7</sub> = 10,39; K<sub>8</sub> = 8,56; K<sub>9</sub> = 6,97; K<sub>10</sub> = 12210; K<sub>11</sub> = 9263; K<sub>12</sub> = 12467. Sehingga koefisien parameter model regresi *spline* ordo dua yang dihasilkan pada Tabel 5.

**Tabel 5.** Koefisien parameter model regresi *spline* kuadratik 12 knot

Parameter	Estimasi	Parameter	Estimasi
$\beta_0$	22.47712204434285	$\beta_{11}$	1.93063986140461
$\beta_1$	-0.01051604557613	$\beta_{12}$	-0.04105331105761
$\beta_2$	0.01284517841570	$\beta_{13}$	-510.31730251373417
$\beta_3$	-0.1452353400126	$\beta_{14}$	931.74610400285264
$\beta_4$	4.7502221611548	$\beta_{15}$	-553.47385007387318
$\beta_5$	-11.443630015824	$\beta_{16}$	0.00112511453235
$\beta_6$	0.86544646857162	$\beta_{17}$	0.00001114201909

$\beta_7$	0.01877889991672	$\beta_{18}$	-0.00001114299410
$\beta_8$	-4.6028024765276	$\beta_{19}$	-0.00001115385600
$\beta_9$	15.8155848526702	$\beta_{20}$	0.00001114241527
$\beta_{10}$	-11.1889134260165		

Berdasarkan Tabel 5, maka koefisien parameter dapat dibuat estimasi model regresi *spline* ordo dua terbaik menggunakan 4 variabel prediktor dan kombinasi 12 titik knot. Estimasi model regresi *spline* ordo dua yang dimaksud adalah sebagai berikut.

$$\begin{aligned} \hat{Y} = & 22,4771 - 0,0105X_1 + 0,01285X_1^2 - 0,14523(X_1 - 76,6)_+^2 \\ & + 4,7502(X_1 - 75,78)_+^2 - 11,4436(X_1 - 73,40)_+^2 + 0,8654X_2 \\ & + 0,01877X_2^2 - 4,6028(X_2 - 13,5)_+^2 + 15,8156(X_2 - 13,05)_+^2 \\ & - 11,1889(X_2 - 13,31)_+^2 + 1,93064X_3 - 0,04105X_3^2 \\ & - 610,3174(X_3 - 10,39)_+^2 + 931,7461(X_3 - 10,42)_+^2 \\ & - 553,4739(X_3 - 10,53)_+^2 + 0,001125X_4 + 1,114202 \times 10^{-5}X_4^2 \\ & - 1,111430 \times 10^{-5}(X_4 - 12210)_+^2 + 1,114242 \times 10^{-5} \\ & - 1,115339 \times 10^{-5}(X_4 - 9263)_+^2 \end{aligned}$$

Berdasarkan estimasi model regresi *spline* ordo dua dapat dimunculkan pula sebagai fungsi sepenggal. sebagai berikut:

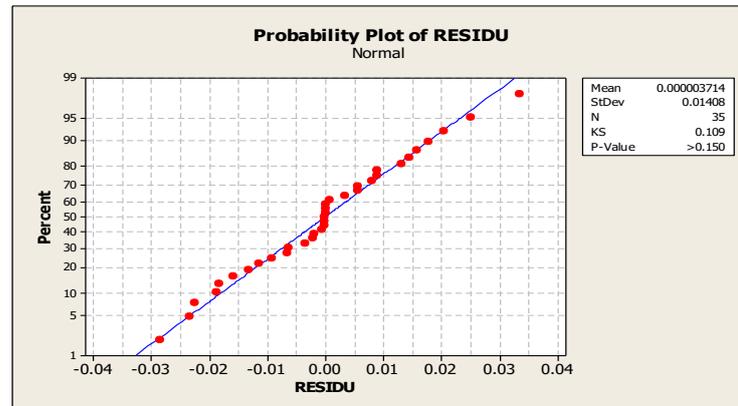
$$\hat{Y} = \left\{ \begin{array}{l} 22,4771 - 0,0105X_1 + 0,0129X_1^2, 73,40 \leq X_1 \leq 76,6 \\ -303,2383 + 8,581X_1 - 0,0534X_1^2, 76,6 \leq X_1 \leq 76,85 \\ 28768,7266 - 743,2154X_1 + 4,8069X_1^2, 76,85 \leq X_1 \leq 77,18 \\ -40578,2176 + 1047,07623X_1 - 6,7478X_1^2, 77,18 \leq X_1 \leq 77,47 \\ 22,4771 + 0,8654X_2 + 0,0188X_2^2, 12,45 \leq X_2 \leq 13,05 \\ -782,0729 + 122,6695X_2 - 4,5942X_2^2, 13,05 \leq X_2 \leq 13,31 \\ 2035,8953 - 300,7667X_2 + 11,3125X_2^2, 13,31 \leq X_2 \leq 13,45 \\ 46,5431 - 2,51303X_2 + 0,1335X_2^2, 13,45 \leq X_2 \leq 13,54 \\ 22,4771 + 1,8208X_3 - 0,03206X_3^2, 6,42 \leq X_3 \leq 6,52 \\ -65861,4582 + 12684,2166X_3 - 610,3495X_3^2, 6,52 \leq X_3 \leq 6,97 \\ 36401,9954 - 6944,0854X_3 + 331,5076X_3^2, 6,97 \leq X_3 \leq 10,39 \\ -23871,1942 + 4503,8138X_3 - 212,0773X_3^2, 10,39 \leq X_3 \leq 10,42 \\ 22,4771 + 0,010125X_4 + 1,114202 \times 10^{-5}X_4^2, 9246 \leq X_4 \leq 9263 \\ 14,2882 + 0,00213X_4 - 5,483322 \times 10^{-8}X_4^2, 9263 \leq X_4 \leq 12467 \\ 22,4519 + 0,000819X_4 - 1,114242 \times 10^{-5}X_4^2, 12467 \leq X_4 \leq 12999 \end{array} \right.$$

Dari fungsi sepenggal di atas menerangkan bahwa setiap titik knot tertentu mempunyai persamaan tertentu pula. Hal ini menjelaskan bahwa model persamaan regresi *spline* ordo dua bersifat fleksibel tergantung dari kombinasi titik knot optimum yang digunakan. Residual untuk estimasi model regresi *spline* ordo dua dapat dilihat pada Tabel 6.

**Tabel 6.** Nilai estimasi dan nilai residual

No	$Y$	$\hat{Y}$	Residual	No	$Y$	$\hat{Y}$	Residual
1	69,95	69,93163	-0,01837	19	75,00	75,01788	0,01788
2	71,98	71,99310	0,0131	20	71,99	71,99818	0,00818
3	68,97	68,95692	-0,01308	21	72,22	72,21337	-0,00663
4	67,45	67,44812	-0,00188	22	74,10	74,10888	0,00888
5	69,81	69,81575	0,00575	23	69,57	69,57576	0,00576
6	72,68	72,67938	-0,00062	24	72,29	72,29890	0,0089
7	68,22	68,25360	0,0336	25	68,65	68,63117	-0,01883
8	69,87	69,87070	0,0007	26	69,63	69,60654	-0,02346
9	74,25	74,25344	0,00344	27	66,32	66,33584	0,01584
10	75,56	75,55077	-0,00923	28	68,39	68,36742	-0,02258
11	76,98	76,97995	-0,00005	29	66,11	66,13052	0,02052
12	70,25	70,27517	0,02517	30	78,99	78,99012	0,00012
13	75,86	75,86007	0,00007	31	82,21	82,20995	-0,00005
14	73,95	73,94656	-0,00344	32	83,14	83,14004	0,00004
15	69,87	69,85869	-0,01131	33	83,05	83,04997	-0,00003
16	68,84	68,81157	-0,02843	34	74,98	74,96419	-0,01581
17	70,02	70,01365	-0,00635	35	75,07	75,0678	-0,0022
18	71,77	71,78453	0,01453				

Pengujian model regresi *spline* perlu dilakukan, agar dapat melihat apakah model regresi *spline* ini sudah baik atau belum. Pengujian yang dilakukan yaitu antaranya uji residual, uji parameter regresi secara serentak dan uji secara individu. Pengujian kenormalan residual digunakan untuk menguji apakah *error* atau model regresi berdistribusi normal atau tidak. Pengujian kenormalan menggunakan Uji Kolmogorov-Smirnov dengan bantuan *software* Minitab 16. Berikut ini merupakan plot kenormalan residual yang diperoleh



**Gambar 2.** Kurva kenormalan residual regresi *spline* kuadrat.

Berdasarkan Gambar 2, dapat dilihat bahwa grafik kenormalan residual menyebar di garis linier, sehingga model regresi *spline* kuadrat yang diperoleh telah memenuhi asumsi residual berdistribusi normal. Pada uji parameter secara serentak dapat digunakan dengan uji  $F$  dengan membandingkan  $F_{hitung}$  dengan  $F_{tabel}$ . Maksud dari pengujian ini yaitu untuk menguji apakah semua variabel prediktor berpengaruh secara serentak terhadap variabel respon. Berikut hipotesis yang digunakan:

**Tabel 7.** Analisis Variansi

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	664,714	166,179	5724,26	0,000
X1	1	11,494	11,494	395,94	0,000
X2	1	4,209	4,209	144,99	0,000
X3	1	9,127	9,127	314,40	0,000
X4	1	25,991	25,991	895,31	0,000
Error	30	0,871	0,029		
Total	34	665,585			

$$F_{hitung} = 5724,26$$

$$F_{tabel (0,05;4;18)} = 2,93$$

Keputusan yang dapat diambil karena nilai  $F_{hitung} > F_{tabel (0,05;2;23)} = 5724,26 > 2,93$ , maka  $H_0$  ditolak artinya variabel  $X_1$ ,  $X_2$ ,  $X_3$  dan  $X_4$  secara bersama-sama berpengaruh secara signifikansi terhadap variabel  $Y$ . Pada uji parameter secara individu dapat dilakukan dengan membandingkan nilai  $t_{hitung}$  dengan nilai  $t_{tabel}$

untuk masing-masing parameter variabel prediktor. Maksud dari pengujian ini yaitu untuk menguji apakah masing masing variabel prediktor berpengaruh secara signifikan terhadap variabel respon.

**Tabel 8.** Nilai *t* hitung parameter regresi

Parameter Regresi	<i>t</i> hitung	Parameter Regresi	<i>t</i> hitung
$\beta_0$	3.033229	$\beta_{10}$	2.60757
$\beta_3$	-109.616	$\beta_{13}$	4.524787
$\beta_8$	-59.1343	$\beta_{14}$	-4.67471
$\beta_9$	-2.8306	$\beta_{15}$	-4.68628

$$t_{tabel}(\alpha/2, n-q-M-1) = t_{tabel}(0,025;18) = 2,10092$$

Pada  $\beta_0, \beta_3, \beta_8, \beta_9, \beta_{10}, \beta_{13}, \beta_{14}, \beta_{15}$  nilai  $|t_{hitung}| > t_{tabel}$ , maka  $H_0$  ditolak. Kesimpulan yang dapat diperoleh masing-masing berpengaruh secara signifikan terhadap model.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, maka dapat diambil kesimpulan:

1. Model regresi *spline* kuadratik terbaik diperoleh dari kombinasi 12 titik knot optimum yang digunakan adalah  $K_1 = 76,6; K_2 = 75,78; K_3 = 73,40; K_4 = 13,5; K_5 = 13,05; K_6 = 13,31; K_7 = 10,39; K_8 = 8,56; K_9 = 6,97; K_{10} = 12210; K_{11} = 9263; K_{12} = 12467$ . Model regresi *spline* ordo dua terbaik pada faktor-faktor yang mempengaruhi presentase indeks pembangunan manusia di Provinsi Jawa Tengah Tahun 2020 yaitu presentase umur harapan hidup saat lahir ( $X_1$ ), harapan lama sekolah ( $X_2$ ), rata-rata lama sekolah ( $X_3$ ), dan pengeluaran per kapita ( $X_4$ ).

2. Estimasi model adalah sebagai berikut:

$$\begin{aligned} \hat{Y} = & 22,4771 - 0,0105X_1 + 0,01285X_1^2 - 0,14523(X_1 - 76,6)_+^2 \\ & + 4,7502(X_1 - 75,78)_+^2 - 11,4436(X_1 - 73,40)_+^2 + 0,8654X_2 \\ & + 0,01877X_2^2 - 4,6028(X_2 - 13,5)_+^2 + 15,8156(X_2 - 13,05)_+^2 \\ & - 11,1889(X_2 - 13,31)_+^2 + 1,93064X_3 - 0,04105X_3^2 \\ & - 610,3174(X_3 - 10,39)_+^2 + 931,7461(X_3 - 10,42)_+^2 \\ & - 553,4739(X_3 - 10,53)_+^2 + 0,001125X_4 + 1,114202 \times 10^{-5}X_4^2 \\ & - 1,111430 \times 10^{-5}(X_4 - 12210)_+^2 + 1,114242 \times 10^{-5} - 1,115339 \\ & \times 10^{-5}(X_4 - 9263)_+^2 \end{aligned}$$

Model hanya berlaku pada interval tertentu, yaitu  $73,40 \leq X_1 \leq 76,6$ ;  $13,05 \leq X_2 \leq 13,5$ ;  $6,97 \leq X_3 \leq 10,39$  dan  $9263 \leq X_4 \leq 12467$ . Model tersebut memperoleh nilai minimum MSE sebesar 0,01181046, dengan nilai koefisien determinasi sebesar 98,8861%. Saran dapat dilakukan perhitungan dengan memperbanyak variabel bebas dan kombinasi titik knot, untuk dapat membandingkan model yang terbaik.

$\beta_0$	205,1145008	236,880186	16,7180015	-211,5870
$\beta_1$	-4,0310337	-38,880438	10,3086680	0,07165321
$\beta_2$	0,0303780	1,112513	-0,3264164	-0,000003257513
$\beta_3$	5,7506581	-605,430042	2501,1763070	-0,00000001747584
$\beta_4$	-507,8600806	1552,622065	-6201,2135730	0,000001140067
$\beta_5$	1320,0285553	-1852,146613	3231,1787562	0,000001128648

#### DAFTAR PUSTAKA

- Beniar, B., *Model Regresi Spline Kuadrat Multivariabel Pada Gross Domestik Product Indonesia*, Jurusan FMIPA, Fakultas Matematika dan Ilmu Pengetahuan Alam, UNSOED, 2016.
- Eubank, R., *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1998.
- Hardle, W., *Applied Nonparametric Regression*, Cambridge University Press, New York, 1990.
- .Nurmalisa, W. dan Nur, I. M., *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*, Prosiding Mahasiswa Seminar Nasional Jurusan Statistika UNIMUS, Semarang, 2019.
- Ruppert, D., Wand, M. P., dan Carroll, R. J., *Semiparametric Regression*, Cambridge University Press, New York, 2003.

Tripena, A., *Pendekatan Model Regresi Spline Linier*. Jurusan MIPA, Fakultas Sains dan Teknik, UNSOED, 2005.

Tripena, A., *Analisis Regresi Spline Kuadrat*, Prosiding Seminar Nasional Matematika dan Pendidikan Matematika Jurusan Pendidikan Matematika, FMIPA UNY, Yogyakarta, 2011.

