



# PEMODELAN REGRESI NONPARAMETRIK SPLINE DAN APLIKASINYA PADA INDEKS KEBAHAGIAAN PROVINSI DI INDONESIA

AGNES ESTER TETRAPOIK<sup>1</sup>, FERRY KONDO LEMBANG<sup>2\*</sup>, VENN YAN ISHAK ILWARU<sup>3</sup>,  
NORISCA LEWAHERILLA<sup>4</sup>

<sup>1</sup>Jurusan Matematika, FMIPA Universitas Pattimura, Indonesia

\*ferrykondolembang@gmail.com

## ABSTRAK

Indeks Kebahagiaan saat ini menjadi salah satu indikator kesejahteraan yang menjadi fokus utama pemangku kepentingan. Di Indonesia, Indeks Kebahagiaan hanya terdapat sampai pada level provinsi dan memiliki perbedaan. Perbedaan Indeks Kebahagiaan setiap Provinsi dipengaruhi oleh banyak faktor sehingga tujuan dari penelitian ini ingin dianalisis faktor-faktor yang mempengaruhi Indeks Kebahagiaan Provinsi di Indonesia. Hasil penelitian menunjukkan model Regresi Nonparametrik Spline terbaik Indeks Kebahagiaan Provinsi di Indonesia yaitu model dengan kombinasi titik knot 2,1,3,2,3 yang memiliki nilai *Generalized Cross Validation* (GCV) minimum sebesar 2,38 dengan empat variabel prediktor yang berpengaruh signifikan antara lain Angka Melek Huruf, Tingkat Pengangguran Terbuka, Perilaku Hidup Bersih dan Sehat, dan Persentase Penduduk Miskin.

**Kata Kunci:** Indeks Kebahagiaan, Regresi Nonparametrik Spline, GCV, Titik Knot

## ABSTRACT

*The Happiness Index is currently one of the indicators of well-being that is the main focus of stakeholders. In Indonesia, the Happiness Index is only found at the provincial level and has differences. The difference in the Happiness Index of each Province is influenced by many factors so that the purpose of this study is to analyze the factors that affect the Happiness Index of Provinces in Indonesia. The results showed that the best Spline Nonparametric Regression model for the Provincial Happiness Index in Indonesia was a model with a combination of 2,1,3,2,3 knot points which had a minimum Generalized Cross Validation (GCV) value of 2.38 with four predictor variables that had a significant effect between Others are Literacy Rate, Open Unemployment Rate, Clean and Healthy Life Behavior, and Percentage of Poor Population.*

**Keywords:** *The Happiness Index, Spline Nonparametric Regression, GCV, Knot Point.*

## 1 Pendahuluan

Kesejahteraan masyarakat sangatlah penting bagi negara karena menjadi salah satu indikator penentu pertumbuhan ekonomi. Dasar-dasar ideologis dari kesejahteraan negara modern adalah keyakinan bahwa seseorang dapat dibuat lebih bahagia dengan memberikan kondisi kehidupan yang lebih baik [1]. Oleh sebab itu, kesejahteraan juga bisa diukur berdasarkan tingkat kebahagiaan dari suatu negara. Salah satu pendekatan yang digunakan untuk mengukur kebahagiaan penduduk di Indonesia yaitu melalui Indeks Kebahagiaan (*Happiness Index*). Definisi Indeks Kebahagiaan sendiri menurut Badan Pusat Statistik (BPS) yakni tingkat kebahagiaan/kepuasan hidup penduduk Indonesia dalam skala 0-100.

Indonesia berperingkat ke-92 dari 156 negara dalam daftar negara bahagia di dunia versi *World Happiness Report* (WHR) 2019, dengan total poin sebesar 5.192. Indikasi poin kebahagiaan yang dirilis oleh WHR tersebut disusun didasari oleh sejumlah faktor, antara lain adalah harapan hidup, dukungan sosial, serta tingkat korupsi. Di Indonesia sendiri melalui metode survei, pengukuran terhadap tingkat kebahagiaan sudah dilakukan sebanyak dua kali yaitu di tahun 2014 dan tahun 2017 oleh BPS. Indeks Kebahagiaan penduduk yang diukur pada tahun 2014 memakai dimensi kepuasan hidup, sementara untuk tahun 2017 melalui instrument survey yang dikembangkan berdasarkan ukuran kondisi objektif dan tingkat kesejahteraan subjektif meliputi tiga dimensi besar kebahagiaan seperti Kepuasan Hidup, Perasaan atau kondisi emosional (*affect*), dan makna hidup (*eudaimonia*).

Penentuan faktor-faktor apa saja yang dapat mempengaruhi Indeks Kebahagiaan dapat dilakukan dengan berbagai metode dimana satu diantaranya yaitu analisis regresi. Analisis regresi adalah suatu metode statistika yang digunakan untuk mengukur ada atau tidaknya hubungan antara satu atau lebih variabel. Terdapat 3 jenis pola kurva dalam regresi yakni parametrik, nonparametrik, dan semiparametrik. Apabila kurva regresi membentuk pola tertentu seperti kubik, linier, atau kuadrat maka digunakan metode regresi parametrik. Regresi nonparametrik dipakai apabila kurva regresi tidak membentuk pola tertentu sedangkan regresi semiparametrik digunakan jika dalam suatu model regresi ditemui komponen parametrik dan nonparametrik. Dalam kenyataan yang ada, pola regresi yang diperoleh tidak selalu membentuk tertentu (regresi parametrik). Oleh sebab itu, apabila kurva regresinya tidak membentuk pola maka metode yang tepat digunakan yakni metode regresi nonparametrik dan semiparametrik.

Secara matematis dan logika individual dapat dikatakan bahwa variabel prediktor yang mempengaruhi Indeks Kebahagiaan memiliki pola data yang linier. Pola data yang linier sebab secara sistematis dan logika apabila variabel prediktor seperti perilaku hidup bersih dan sehat semakin tinggi maka Indeks Kebahagiaan semakin tinggi. Akan tetapi, teori ini tidak bisa dipakai pada penelitian ini sebab unit observasi penelitian adalah Provinsi di Indonesia dimana tentunya diketemukan variasi dalam data. Oleh sebab itu, bentuk pola data akan acak dan tidak membentuk pola, sehingga metode analisis yang tepat yaitu regresi nonparametrik spline. Kekhasan penggunaan metode regresi nonparametrik spline cenderung disebabkan oleh spline dimana merupakan potongan-potongan polinomial tersegmen yang memiliki sifat fleksibilitas tinggi yang bisa menyesuaikan diri dengan karakteristik data [2]

Beberapa penelitian terdahulu yang pernah dilakukan menggunakan metode Regresi Nonparametrik Spline antara lain yang dilakukan oleh [3] dimana menerapkan pada data Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar dengan kesimpulan model terbaik adalah tiga titik knot. Selanjutnya, penelitian oleh [4] dalam memodelkan persentase penduduk miskin di Kalimantan. Penelitian dengan metode yang sama dilakukan oleh [5] yakni menganalisis data Indeks Pembangunan Manusia (IPM) di Indonesia Tahun 2018 dan penelitian yang dilakukan oleh [6] dalam memodelkan angka morbiditas di Indonesia.

Merujuk relevansi penelitian sejenis maka peneliti tertarik melakukan penelitian dengan alat analisis statistik yang sama pada data Indeks Kebahagiaan Provinsi di Indonesia dengan menambahkan model kombinasi titik knot. Penelitian ini bertujuan untuk memodelkan data Indeks Kebahagiaan Provinsi di Indonesia Tahun 2017 serta ingin diketahui faktor-faktor yang signifikan mempengaruhi Indeks Kebahagiaan Provinsi di Indonesia Tahun 2017 menggunakan metode regresi nonparametrik spline. Hasil penelitian ini diharapkan dapat diketahui faktor-faktor yang berpengaruh signifikan terhadap Indeks Kebahagiaan Provinsi di Indonesia tahun 2017 harapannya dapat memperkaya informasi kepada pihak pengambil kebijakan dalam merumuskan sejumlah kebijakan-kebijakan strategis yang ada dalam rangka meningkatkan kebahagiaan masyarakat.

## 2 Tinjauan Pustaka

### 2.1 Regresi Nonparametrik Spline

Pendekatan regresi nonparametrik umumnya digunakan jika pola data tidak membentuk pola tertentu misalnya pola linier, kuadrat, kubik, dan lain sebagainya. Oleh sebab itu, tanpa dipengaruhi subjektivitas dari peneliti diharapkan data bisa menemukan sendiri bentuk estimasi kurva regresi atau sebutan lainnya, regresi nonparametrik mempunyai fleksibilitas yang tinggi [7]. Regresi spline memiliki titik knot yang merupakan suatu titik perpaduan yang terbentuk karena perubahan pola perilaku kurva pada interval yang berbeda [8]. Berikut adalah persamaan matematis dari model regresi nonparametrik.

$$y_i = f(x_i) + \varepsilon_i \quad ; i = 1, 2, 3, \dots, n \quad (1)$$

Lain halnya dengan regresi parametrik, regresi nonparametrik umumnya digunakan untuk estimasi kurva regresi sebab model tidak ditentukan terlebih dahulu.

### 2.2 Pemilihan Titik Knot Optimal

Ada berbagai metode yang bisa dipakai ketika memilih titik *knot* optimal diantaranya yaitu metode GCV (*Generalized Cross Validation*). Keunggulan menggunakan metode GCV antara lain karena memiliki sifat optimal asimtotik dimana tidak dimiliki oleh metode lain semisal metode *Cross Validation* (CV) dan dapat dipakai untuk sampel besar. Model Regresi spline terbaik didapatkan dari titik knot optimal berdasarkan nilai GCV minimum. Berikut formula yang dipakai untuk menghitung nilai GCV [9].

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

Dimana  $\mathbf{I}$  merupakan matriks identitas,  $n$  yaitu jumlah observasi,  $\mathbf{A}(K_1, K_2, \dots, K_j)$  adalah matriks  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  sedangkan

$$MSE(K_1, K_2, \dots, K_j) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

### 2.3 Koefisien Determinasi

Koefisien Determinasi merupakan alat ukur proporsi keragaman atau variansi total disekeliling nilai tengah (mean)  $y$  yang mampu diterangkan oleh model regresi. Idealnya, semakin tinggi nilai koefisien determinasi maka semakin baik juga model yang diperoleh sebab mampu menerangkan keragaman variabel respon [10]. Formula Koefisien Determinasi dapat ditulis sebagai berikut:

$$R^2 = \frac{SS_{regresi}}{SS_{total}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

### 2.4 Pengujian Signifikansi Parameter Model Secara Serentak

Tidak jauh berbeda dengan regresi parametrik, tujuan dilakukannya pengujian signifikansi parameter model secara serentak pada model regresi nonparametrik spline yakni untuk mengetahui seluruh parameter secara bersama-sama signifikan terhadap variabel respon atau tidak [11]. Adapun hipotesis yang digunakan dalam pengujian serentak sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{m+r} = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0 \quad ; j = 1, 2, \dots, m+r$$

Statistik uji untuk pengujian signifikansi serentak [12]:

$$F_{hitung} = \frac{MS_{regresi}}{MS_{error}} \quad (4)$$

Daerah Kritis : Tolak  $H_0$  jika  $F_{hitung} > F_{tabel}(F_{\alpha;(m+r,n-(m+r)-1})$  atau p-value  $< \alpha$

## 2.5 Pengujian Signifikansi Parameter Model Secara Parsial

Tujuan dilakukannya pengujian signifikansi parameter model secara parsial untuk mengetahui seberapa jauh variabel prediktor secara individu dalam menerangkan variasi variabel respon. Adapun hipotesis yang digunakan dalam pengujian parsial sebagai berikut:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad ; j = 1, 2, \dots, m+r$$

Statistik uji yang dipakai untuk pengujian secara parsial yaitu uji  $t$  [13] dengan formula sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (5)$$

Daerah kritis : Tolak  $H_0$  jika  $|t_{hitung}| > t_{\alpha/2, n-(m+r)-1}$  atau p-value  $< \alpha$

## 2.6 Uji asumsi identik

Model yang dihasilkan dalam analisis regresi nonparametrik spline *Truncated* harus memenuhi asumsi residual identik. Tujuan dilakukannya uji identik dalam pemodelan regresi nonparametrik spline antara lain untuk melihat homogenitas dari variansi residual. Menurut [14] dikatakan bahwa asumsi identik terpenuhi jika variansi antar residual sama atau dengan kata lain yakni tidak terjadi heterokedastisitas. Salah satu alat analisis statistik yang digunakan untuk melakukan pengujian asumsi identik yakni Uji Glejser dengan hipotesis sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (Residual Identik)}$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, n \text{ (Residual Tidak Identik)}$$

Statistik uji:

$$F_{hitung} = \frac{MS_{regresi}}{MS_{residual}} = \frac{\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}|)^2 / k - 1}{\sum_{i=1}^n (|e_i| - |\bar{e}|)^2 / n - k} \quad (6)$$

Daerah Kritis : Tolak  $H_0$  apabila  $F_{hitung} > F_{(\alpha;1;n-2)}$  atau p-value  $< \alpha$

## 2.7 Uji asumsi distribusi normal

Selain uji asumsi identik, salah satu asumsi yang wajib dipenuhi dalam pemodelan regresi nonparametrik spline yaitu residual model berdistribusi normal. Tujuan utama dilakukannya pengujian asumsi distribusi normal yakni untuk melihat apakah residual telah mengikuti distribusi normal atau tidak. Umumnya, alat analisis statistik yang digunakan untuk melakukan pengujian asumsi distribusi normal yaitu uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut.

$$H_0 : F(x) = F_0(x) \text{ (Residual Mengikuti Distribusi Normal)}$$

$$H_1 : F(x) \neq F_0(x) \text{ (Residual Tidak Mengikuti Distribusi Normal)}$$

Statistik uji:

$$D = \sup_x |S(x) - F_0(x)| \quad (7)$$

Daerah Kritis : Tolak  $H_0$  apabila  $D < D_{(1-\alpha,n)}$  [15]

### 3. Metodologi Penelitian

#### 3.1 Data

Penelitian ini menggunakan data sekunder yang diperoleh dari sejumlah rilis yang dikeluarkan oleh BPS pada website <https://www.bps.go.id/>. Data tersebut terdiri atas 5 variabel prediktor (X) dan 1 variabel respon (Y) yakni Indeks Kebahagiaan (Y), Angka Melek Huruf ( $X_1$ ), Tingkat Pengangguran Terbuka ( $X_2$ ), Kebijakan Perilaku Hidup Bersih dan Sehat ( $X_3$ ), Kepadatan Penduduk ( $X_4$ ), Persentase Penduduk Miskin ( $X_5$ ). Unit penelitian yang digunakan yaitu seluruh Provinsi di Indonesia yaitu sejumlah 34 Provinsi Pada Tahun 2017.

#### 3.2 Prosedur Analisis Data

Analisis data pada penelitian ini menggunakan bantuan software ARCGIS, R Studio Version 1.4.1717 dan Minitab 17. Tahapan analisis pada penelitian ini diuraikan sebagai berikut :

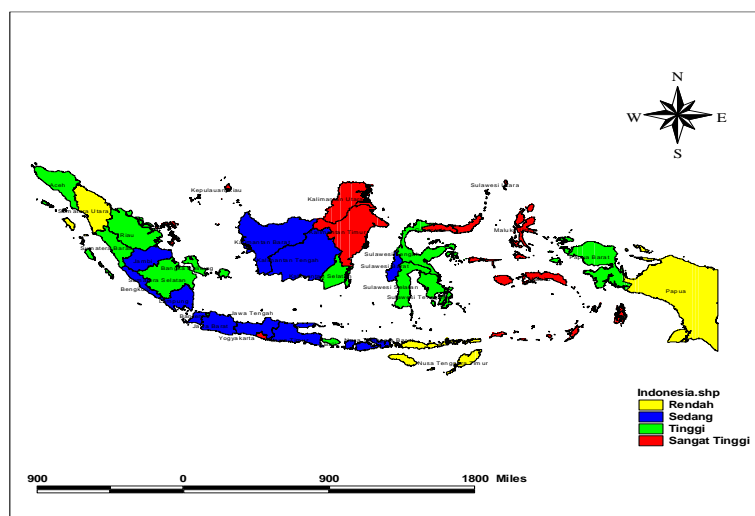
1. Mendeskripsikan Variabel Respon (Y) secara visual melalui peta tematik.
2. Melakukan langkah-langkah pemodelan Regresi Nonparametrik *Spline* seperti berikut:
  - a. Membuat *scatterplot* antara variabel respon dengan setiap variabel prediktor untuk melihat pola hubungan.
  - b. Menentukan titik knot optimal antara satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot berdasarkan nilai GCV minimum sebagai dasar pemilihan model *spline* terbaik.
  - c. Model *spline* terbaik diuji signifikansi parameter secara simultan menggunakan uji F dan secara parsial menggunakan uji t.
  - d. Melakukan pengujian kelayakan model melalui Uji Asumsi Residual antara lain pengujian asumsi identik menggunakan Uji Glejser dan distribusi normal menggunakan Uji Kolmogorov-Smirnov pada residual data.

### 4. Hasil dan Pembahasan

Tahapan analisis dalam penelitian ini mengikuti seluruh kajian teori hingga memperoleh hasil dimana dituangkan kedalam beberapa point dibawah ini.

#### 4.1. Deskriptif Indeks Kebahagiaan Provinsi di Indonesia

Deskriptif data untuk memberikan gambaran data agar informasi yang disajikan lebih mudah dipahami. Adapun deskripsi data Indeks Kebahagiaan Provinsi di Indonesia ditampilkan melalui peta tematik dibawah ini.



**Gambar 1:** Peta Sebaran Indeks Kebahagiaan Provinsi di Indonesia

Berdasarkan Gambar 1, nilai Indeks Kebahagiaan di Indonesia dibagi atas 4 kategori yaitu kategori rendah, sedang, tinggi dan sangat tinggi. Provinsi di Indonesia dengan nilai

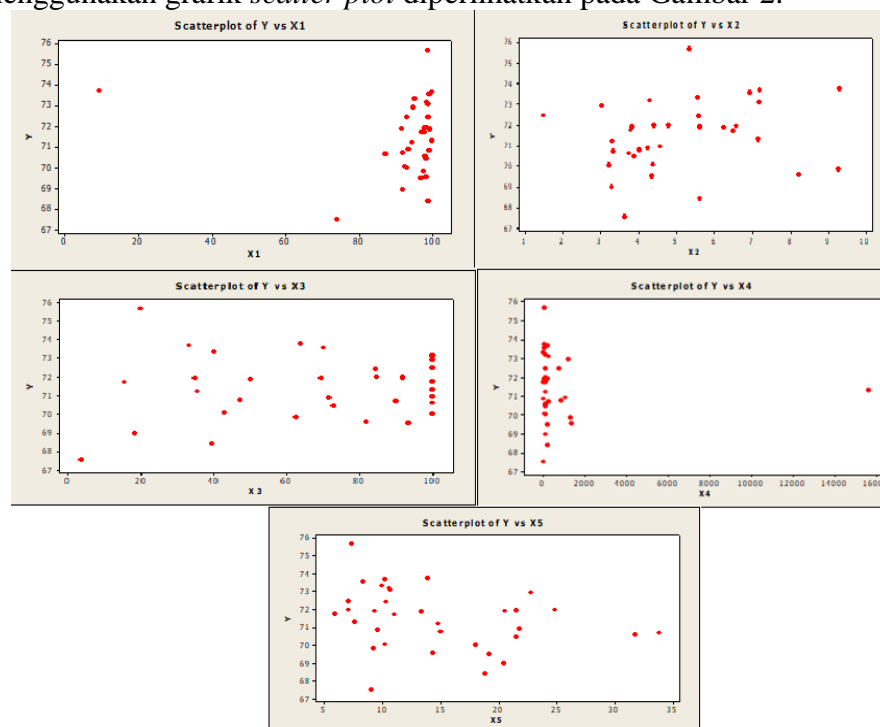
Indeks Kebahagiaan berkategori rendah antara lain Provinsi Sumatera Utara, Nusa Tenggara Timur dan Papua. Untuk kategori sedang yaitu Provinsi Jambi, Bengkulu, Lampung, Banten, Jawa Barat, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Sulawesi Barat. Provinsi Aceh, Riau, Sumatera Barat, Bangka Belitung, Sumatera Selatan, Kalimantan Selatan, DI Yogyakarta, Bali, Sulawesi Tengah, DKI Jakarta, Sulawesi Tenggara, Sulawesi Selatan, Papua Barat termasuk provinsi dengan Indeks Kebahagiaan berkategori tinggi sedangkan Provinsi Kalimantan Timur, Kalimantan Utara, Kep. Riau, Gorontalo, Sulawesi Utara, Maluku Utara dan Maluku merupakan provinsi dengan nilai Indeks Kebahagiaan paling tinggi.

#### 4.2. Pemodelan Regresi Nonparametrik Spline

Tahapan selanjutnya yang dilakukan dalam penelitian ini yakni menerapkan model Regresi Nonparametrik Spline untuk mengetahui variabel prediktor mana saja yang berpengaruh signifikan terhadap indeks kebahagiaan provinsi di Indonesia tahun 2017. Akan tetapi, sebelum penerapan dimaksud dilakukan terlebih dahulu diidentifikasi pola hubungan dari setiap variabel prediktor dengan variabel respon menggunakan grafik scatter plot sebagai sebuah penegasan terhadap ketepatan penggunaan model Regresi Nonparametrik Spline dalam kasus ini.

##### a. *Scatter Plot* Indeks Kebahagiaan dan Faktor-Faktor Yang Mempengaruhinya.

Identifikasi pola hubungan antara indeks kebahagiaan dengan faktor-faktor yang diduga mempengaruhi menggunakan grafik *scatter plot* diperlihatkan pada Gambar 2.



**Gambar 2:** *Scatter plot* antara Indeks Kebahagiaan dan Variabel Prediktor

Berdasarkan Gambar 2 di atas diperoleh informasi bahwa *Scatter plot* antara indeks kebahagiaan dengan variabel-variabel yang diduga mempengaruhi tidak memperlihatkan adanya kecenderungan membentuk pola tertentu oleh karena itu model regresi nonparametrik spline dapat digunakan untuk pemodelan selanjutnya.

##### b. Titik Knot Optimal

Perbandingan nilai GCV minimum dalam menentukan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot ditunjukkan pada Tabel 3 dibawah ini

**Tabel 3 : Perbandingan nilai GCV**

No.	Titik Knot	GCV
1.	Satu Titik Knot	2,747029
2.	Dua Titik Knot	5,583999
3.	Tiga Titik Knot	2,372454
4.	Kombinasi Titik Knot 2,1,3,2,3	<b>2,371608</b>

Berdasarkan Tabel 3 diatas terlihat dengan jelas bahwa nilai GCV terkecil dihasilkan oleh pemilihan kombinasi titik knot 2,1,3,2,3 yakni sebesar 2,371608. Hal ini dapat diartikan bahwa kombinasi titik knot menjadi pilihan terbaik untuk memodelkan Indeks Kebahagiaan beserta faktor-faktor yang diduga mempengaruhi. Berikut ini ditunjukkan persamaan model regresi nonparametrik spline dengan kombinasi titik knot

$$\hat{y} = 10,63366 + 6,802687x_1 - 20,765(x_1 - 49,41)_+^1 + 13,99384(x_1 - 69,55)_+^1 + 0,227493x_2 - 1,87065(x_2 - 6,86667)_+^1 + 0,315282x_3 - 0,49545(x_3 - 24,90556)_+^1 + 0,243843(x_3 - 46,36111)_+^1 - 0,06437(x_3 - 67,81667)_+^1 + 0,000919x_4 - 0,0013(x_4 - 6948,766)_+^1 - 0,00078(x_4 - 10418,7)_+^1 - 0,03555x_5 - 0,4869(x_5 - 12,08333)_+^1 + 1,007025(x_5 - 18,27667)_+^1 - 0,63272(x_5 - 24,47)_+^1$$

Model regresi nonparametrik spline dengan kombinasi titik knot mempunyai nilai koefisien determinasi ( $R^2$ ) sebesar 73,76%. Itu artinya bahwa model diatas mampu menjelaskan Indeks Kebahagiaan sebesar 73,76%.

Salah satu bentuk interpretasi model regresi nonparametrik spline berdasarkan model terbaik diatas pada masing-masing variabel adalah sebagai berikut.

1. Apabila  $x_2$ ,  $x_3$  dan  $x_5$  diasumsikan konstan maka pengaruh dari variabel Angka Melek Huruf adalah

$$\hat{y} = 6,82687x_1 - 20,765(x_1 - 49,41)_+^1 + 13,99384(x_1 - 69,55)_+^1 = \left\{ \begin{array}{ll} 6,82687x_1 & ; x_1 < 49,41 \\ -13,93813x_1 + 1025,99865 & ; 49,41 \leq x_1 < 69,55 \\ 0,0571x_1 - 973,27 & ; x_1 > 69,55 \end{array} \right\}$$

Berdasarkan model tersebut menjelaskan bahwa ketika suatu provinsi di Indonesia yang memiliki AMH yang kurang dari 49,41 mengalami peningkatan AMH sebesar satuan maka indeks kebahagiaan di Provinsi tersebut akan bertambah sebesar 6,82687. Ketika suatu Provinsi di Indonesia yang mengalami peningkatan pada interval 49,41 sampai 69,55 mengalami peningkatan AMH sebesar satu satuan maka indeks kebahagiaan tersebut akan mengalami penurunan sebesar 13,93813. Ketika suatu Provinsi di Indonesia memiliki nilai AMH yang lebih dari 69,55 mengalami peningkatan sebesar satu satuan maka indeks kebahagiaan di Provinsi tersebut akan bertambah sebesar 0,0571.

**c. Pengujian Signifikansi Model Secara Simultan**

Uji serentak dilakukan pada parameter model regresi terhadap variabel Indeks Kebahagiaan secara simultan. Dibawah ini ditunjukkan hasil uji simultan model regresi nonparametrik spline.

**Tabel 4: Uji Simultan**

Sumber Variasi	Derajat Bebas	Jumlah Kuadrat	Rataan Kuadrat	$F_{hitung}$	$F_{0,05 (16;17)}$
Regresi	16	70,777	4,4236	2,9864	2,29
Error	17	25,181	1,4812		
Total	33	95,958			

Berdasarkan tabel 4 terlihat bahwa diperoleh nilai  $F_{hitung} = 2,9864$ . Dalam pengujian ini digunakan taraf signifikansi sebesar 5% sehingga diperoleh nilai  $F_{0,05 (16;17)} = 2,29$ . Dari perbandingan nilai  $F_{hitung}$  dan nilai  $F_{0,05 (16;17)}$  maka keputusan yang diambil yakni Tolak  $H_0$  artinya minimal terdapat satu parameter model yang signifikan.

#### d. Pengujian Signifikansi Model Secara Parsial

Pengujian parsial bertujuan untuk mengetahui parameter mana yang signifikan terhadap model regresi. Dibawah ini disajikan hasil pengujian signifikansi parameter secara parsial.

**Tabel 5:** Uji Parsial Estimasi Parameter Spline Linier Dengan Kombinasi Titik Knot

Variabel	Parameter	Estimator	p-value	Keterangan
Constant	$\beta_0$	10,63366	$5,53 e^{-16}$	Signifikan
$x_1$	$\beta_1$	6,802687	$1,35 e^{-17}$	Signifikan
	$\beta_2$	-20,765	$2,53 e^{-16}$	Signifikan
	$\beta_3$	13,99384	$6,70 e^{-15}$	Signifikan
$x_2$	$\beta_4$	0,227493	0,371	Tidak Signifikan
	$\beta_5$	-1,87065	0,0408	Signifikan
$x_3$	$\beta_6$	0,315282	0,0300	Signifikan
	$\beta_7$	-0,49545	0,0233	Signifikan
	$\beta_8$	0,243843	0,0640	Tidak Signifikan
	$\beta_9$	-0,06437	0,400	Tidak Signifikan
$x_4$	$\beta_{10}$	0,000919	0,325	Tidak Signifikan
	$\beta_{11}$	-0,0013	0,283	Tidak Signifikan
	$\beta_{12}$	-0,00078	0,283	Tidak Signifikan
$x_5$	$\beta_{13}$	-0,3555	0,867	Tidak Signifikan
	$\beta_{14}$	-0,4869	0,183	Tidak Signifikan
	$\beta_{15}$	1,007025	0,00780	Signifikan
	$\beta_{16}$	-0,63272	0,0845	Tidak Signifikan

Berdasarkan tabel 5 diperoleh informasi bahwa dengan taraf signifikansi sebesar 5%, ada 17 parameter yang terbentuk dimana sebanyak 8 parameter yang signifikan karena memiliki nilai p-value kurang dari taraf signifikansi yang ditetapkan sisanya sebanyak 9 parameter tidak signifikan. Akan tetapi, dalam pemodelan regresi nonparametrik spline dimungkinkan bahwa jika terdapat satu parameter model yang signifikan maka dapat disimpulkan bahwa variabel dimaksud signifikan mempengaruhi variabel respon.

#### e. Uji Asumsi Residual

Untuk mengetahui sejauh mana suatu model regresi layak maka perlu dilakukan uji asumsi residual. Dalam pemodelan regresi nonparametrik spline terdapat dua asumsi yang wajib dipenuhi antara lain yaitu asumsi identik dan distribusi normal. Jika model regresi yang dibentuk belum memenuhi kedua asumsi dimaksud maka dapat disimpulkan bahwa model tersebut belum layak dipakai untuk pemodelan terhadap variabel respon.



Terpenuhinya asumsi residual identik jika kondisi varians residual homogen. Salah satu Uji yang digunakan untuk menguji asumsi residual identik yakni Uji Glejser. Berikut dibawah ini disajikan ANOVA dari Uji Glejser.

**Tabel 6 : ANOVA Uji Glejser**

Model	Derajat Bebas (db)	Jumlah Kuadrat	Rataan Jumlah Kuadrat	F <sub>Hitung</sub>	p-value
Regresi	5	7,118	1,424	2,128	0,093
Residual	27	18,060	0,669		
Total	32	25,178			

Berdasarkan Tabel 6 diperoleh informasi bahwa nilai p-value dari model regresi sebesar 0,093 dimana lebih besar dari taraf signifikansi 5%. Oleh sebab itu, keputusan yang diambil adalah Terima  $H_0$  yakni varians residual homogen. Itu artinya model telah memenuhi asumsi identik.

Uji asumsi yang akan dilakukan selanjutnya yaitu Pengujian asumsi normalitas. Dalam pengujian ini umumnya menggunakan Uji Kolmogorov-Smirnov. Hasil dari Uji Kolmogorov-Smirnov ditunjukkan dibawah ini.

**Tabel 7: Hasil Uji Kolmogorov-Smirnov**

KS	p-value
0,088	0,200

Berdasarkan tabel 7 diatas diperoleh informasi bahwa nilai p-value menggunakan Uji Kolmogorov-Smirnov untuk pengujian asumsi normalitas yakni sebesar 0,200. Hal ini berarti bahwa keputusan yang diambil adalah Terima  $H_0$  atau asumsi residual berdistribusi normal telah terpenuhi.

### 3 Kesimpulan

Berdasarkan analisis hasil dan pembahasan diatas, maka dapat disimpulkan bahwa model regresi nonparametrik spline terbaik dari Indeks Kebahagiaan Provinsi di Indonesia yakni pada kombinasi titik knot 2,1,3,2,3 dimana menghasilkan nilai GCV minimum sebesar 2,371608 dan koefisien determinasi ( $R^2$ ) sebesar 0,7376 yang mengandung arti bahwa kemampuan variabel prediktor menjelaskan variabilitas variabel respon sebesar 73,76% sementara sisanya dijelaskan oleh variabel prediktor lainnya diluar penelitian.

### Daftar Pustaka

- [1] R. Veenhoven, "Correlates of happiness: 7,836 findings from 603 studies in 69 nations: 1911-1994," *Unpublished manuscript, Erasmus University, Rotterdam, The Netherlands*, 1994, doi: 10.1016/0043-1354(91)90046-S.
- [2] W. Härdle, *Applied nonparametric regression*, no. 19. Cambridge university press, 1990.
- [3] W. Sanusi, R. Syam, and R. Adawiyah, "Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar)," *JMathCos (Journal of Mathematics, Computations, and Statistics)*, vol. 2, no. 1, pp. 70–81, 2017, doi: 10.35580/jmathcos.v2i1.12460.
- [4] W. Noviani, S. Sifriyani, and I. Purnamasari, "Pemodelan Regresi Nonparametrik Spline Linear Persentase Penduduk Miskin di Kalimantan," *Jurnal Siger Matematika*, vol. 1, no. 2, pp. 35–41, 2020, doi: 10.23960%2Fjjsm.v1i2.2565.
- [5] W. Alwi and M. Musfirah, "Penerapan Regresi Nonparametrik Spline Dalam Memodelkan Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia (IPM)

- di Indonesia Tahun 2018,” *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 9, no. 2, pp. 112–119, 2021, doi: 10.24252/msa.v9i2.23055.
- [6] Megawati, B. Abapihi, I. Yahya, L. Laome, and Mukhsar, “Pemodelan Angka Morbiditas dengan Menggunakan Regresi Nonparametrik Spline di Indonesia,” in *Seminar Nasional Sains dan Terapan VI*, 2022, vol. 6, pp. 79–84.
- [7] S. B. Loklomin and M. Kaseside, “Analisi Angka Kematian Bayi Kabupaten Halmahera Utara dengan Metode Regresi Nonparametrik Spline Truncated,” *VARIANCE: Journal of Statistics and Its Applications*, vol. 3, no. 1, pp. 1–5, 2021, doi: 10.30598/variancevol3iss1page1-5.
- [8] S. U. R. Sari, “Aplikasi Metode Regresi Nonparametrik Spline Multivariabel Untuk Pemodelan Faktor-faktor yang Mempengaruhi Indeks Pembangunan Gender di Provinsi Jawa Barat,” *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 6, no. 2, 2018, doi: 10.26714/jsunimus.6.2.2018.%25p.
- [9] M. F. Bintariningrum and I. N. Budiantara, “Pemodelan regresi nonparametrik spline truncated dan aplikasinya pada Angka Kelahiran Kasar di Surabaya,” *Jurnal Sains dan Seni ITS*, vol. 3, no. 1, pp. D7–D12, 2014, doi: 10.12962/j23373520.v3i1.6098.
- [10] N. R. Draper and H. Smith, *Applied regression analysis*, vol. 326. John Wiley & Sons, 1998.
- [11] S. C. Chapra and R. P. Canale, “Long-term phenomenological model of phosphorus and oxygen for stratified lakes,” *Water Res*, vol. 25, no. 6, pp. 707–715, 1991, doi: 10.1016/0043-1354(91)90046-S.
- [12] I. P. Merdekawati and I. N. Budiantara, “Pemodelan Regresi Spline Truncated Multivariabel pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah,” *Jurnal Sains dan Seni ITS*, vol. 2, no. 1, pp. D19–D24, 2013, doi: 10.12962/j23373520.v2i1.3035.
- [13] W. R. Smith and K. Draper, “Mobilizing affect: A possible effect of day hospital treatment for chronic psychiatric patients.,” *Int J Partial Hosp*, 1992.
- [14] D. N. Gujarati and D. C. Porter, “Basic econometrics (ed.),” *Singapore: McGraw Hill Book Co*, 2003.
- [15] W. W. Daniel, “Statistika nonparametrik terapan,” *Jakarta: PT Gramedia*, pp. 25–31, 1989.