

REGRESI NONPARAMETRIK BIRESPON SPLINE

Nunung Nurdiani ¹⁾, Nar Herrhyanto ²⁾, Dadan Dasari ³⁾

^{1), 2), 3)} Departemen Pendidikan Matematika FPMIPA UPI

*Surel: nurdiani.nunung@yahoo.com

ABSTRAK: Salah satu ciri majunya suatu negara adalah dengan majunya pembangunan manusia. Pembangunan manusia berperan penting dalam suatu negara. Faktor-faktor yang berpengaruh terhadap pembangunan manusia, diantaranya pendidikan dan ekonomi. Salah satu komponen yang berpengaruh terhadap Indeks Pembangunan Manusia adalah tingkat pendidikan yang diukur dengan Angka Melek Huruf dan Rata-Rata Lama Sekolah. Dalam penelitian ini metode yang digunakan untuk memodelkan Angka Melek Huruf dan Rata-Rata Lama Sekolah adalah regresi nonparametrik birespon spline. Metode ini digunakan karena Spline memiliki kelebihan yakni model akan cenderung mencari estimasinya kemanapun data tersebut akan bergerak. Pemilihan model terbaik berdasarkan nilai *Generalized Cross Validation* (GCV) yang minimum. Model terbaik yang dihasilkan dari penelitian ini adalah model linear dengan satu titik knot.

Kata kunci : Pembangunan Manusia, Rata-Rata Melek Huruf, Rata-Rata Lama Sekolah, Regresi Nonparametrik Birespon Spline, GCV.

ABSTRACT: One feature of progress of a country is the advancement of Human Development. Human Development plays an important role in a country. Factors that influence Human Development, including education and the economy. One of components that affect the Human Development index is measured by the level of education Literacy Rate and Mean Year School. In this research method is used to model Literacy Rate and Mean Year School is nonparametric biresponse spline regression. This method is used because the spline has the advantage that the model will tend to look for the estimate wherever the data is moved. Selecting the best model is based on the minimum *Generalized Cross Validation* (GCV). Best model resulting from tis researc is a linear model with one knot.

Keywords: Human Development, Literacy Rate, Mean Year School, nonparametric biresponse spline regression, GCV.

1. PENDAHULUAN

Salah satu ciri majunya suatu negara adalah dengan majunya pembangunan manusia. Pembangunan manusia berperan penting dalam suatu negara. Faktor-faktor yang berpengaruh terhadap pembangunan manusia, diantaranya pendidikan dan ekonomi. Apabila tingkat ekonomi dan tingkat pendidikan tinggi bisa dikatakan suatu negara telah maju. Salah satu komponen yang berpengaruh terhadap Indeks Pembangunan Manusia (IPM) adalah tingkat pendidikan yang diukur dengan angka melek huruf dan rata-rata lama sekolah. Angka Melek Huruf adalah persentase penduduk usia 15 tahun keatas yang dapat membaca dan menulis huruf latin dan atau huruf lainnya. Rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 15 tahun keatas dalam menjalani pendidikan formal. Di Jawa Barat salah satu provinsi di Indonesia merupakan provinsi yang masih rendah dalam pembangunan manusia contohnya pendidikan. Hal ini perlu adanya peningkatan agar warga Jawa Barat mendapat pendidikan yang tinggi.

Regresi spline adalah regresi dimana kurva regresinya didekati oleh fungsi spline. Fungsi spline adalah fungsi polinomial yang lebih khusus dan memberikan fleksibilitas yang lebih tinggi daripada fungsi polinomial pada umumnya. Dalam regresi nonparametrik spline jika terdapat satu variabel respon dan satu variabel prediktor maka diperoleh regresi nonparametrik spline univariabel. Jika dalam analisis regresi terdapat satu variabel respon dengan variabel prediktor lebih dari satu, maka regresi tersebut dinamakan regresi nonparametrik spline multivariabel. Sedangkan regresi birespon didefinisikan sebagai salah satu model regresi yang memiliki variabel respon yang lebih dari satu buah dan diantara variabel-variabel tersebut terdapat korelasi atau hubungan yang kuat, baik secara logika maupun matematis (Simillia, 2007).

2. TINJAUAN PUSTAKA

A. Analisis Regresi Dalam Statistika

Statistika merupakan salah satu ilmu yang memuat sangat banyak cabang ilmu, diantaranya Time Series (runtun waktu), Proses Stokastik, Probabilitas, Rancangan Percobaan, Analisis Regresi, Reliabilitas, dan lain sebagainya (Budiantara, 2009). Analisis regresi merupakan salah satu bidang statistika yang memainkan peran sangat penting. Analisis regresi umumnya digunakan untuk menyelidiki model pola hubungan fungsional antara satu atau lebih variabel. Disamping itu, analisis regresi sangat bermanfaat untuk peramalan (*forecasting*).

Hal pertama yang semestinya dilakukan untuk memodelkan data adalah apakah variabel-variabel tersebut berkorelasi atau tidak. Apabila secara rasional terjadi korelasi, maka dapat dilakukan pemodelan statistika dengan Analisis Regresi.

B. Regresi Parametrik

Regresi Parametrik adalah pendekatan regresi yang digunakan untuk mengetahui bentuk hubungan antara variabel prediktor (X) dan variabel respon (Y) dengan asumsi bentuk kurva regresi diketahui berdasarkan informasi atau teori sebelumnya. Pendekatan model regresi parametrik memiliki sifat yang sangat baik dari pandangan statistika inferensi (Budiantara, 2009), seperti sederhana, mudah interpretasinya, parsimoni, estimatornya tidak bias, tergolong estimator linear, efisien, konsisten, BLUE (*Best Linear Unbiased Estimator*), yang sangat jarang dimiliki oleh pendekatan model regresi lain seperti regresi nonparametrik dan regresi semiparametrik.

Masalah yang sering muncul dalam regresi adalah tidak semua variabel prediktor (X) dapat didekati dengan pendekatan statistika parametrik, karena tidak adanya informasi tentang bentuk hubungan variabel prediktor (X) dan variabel respon (Y), sehingga harus digunakan pendekatan statistika nonparametrik.

C. Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon yang tidak diketahui bentuk fungsinya. Hal ini karena sebelumnya tidak terdapat informasi tentang bentuk $f(x)$ (Eubank, 1988). Sehingga model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linear ataupun nonlinear. Regresi nonparametrik sangat memperhatikan fleksibilitas, dan hanya diasumsikan bentuk fungsinya mulus (Eubank, 1988). Kekurangan dan kelebihan setiap pemilihan prosedur pengujian data, apakah itu menggunakan nonparametrik atau parametrik memiliki kelebihan dan kekurangan masing-masing.

D. Regresi Semiparametrik

Disamping pendekatan regresi parametrik dan regresi nonparametrik, terdapat pula golongan Statistikawan yang memandang kurva regresi dapat diklasifikasikan kedalam dua komponen, yaitu komponen parametrik (bentuk fungsinya diketahui) dan komponen nonparametrik (bentuk fungsinya tidak diketahui). Pandangan ini memberikan pendekatan regresi semiparametrik (Wahba, 1990; Budiantara, dkk, 2010; Budiantara, 2007; Wu & Zhang 2006). Apabila bagian parametriknya dapat dipolakan linear, maka regresi semiparametrik ini, disebut sebagai regresi linear parsial. Oleh karena itu, pada

regresi semiparametrik penaksiran untuk kurva regresi diperoleh secara ekuivalen dengan penaksiran fungsi dan penaksiran parameter dalam model.

E. Fungsi Spline

Spline merupakan fungsi potongan polinomial yang memiliki sifat tersegmen dan kontinu (*spline polynomial truncated*), dimana dalam fungsi tersebut terdapat titik-titik penghubung yang disebut dengan titik knot. Titik knot adalah titik perpaduan bersama dimana terdapat perubahan pola perilaku fungsi atau kurva. Bentuk umum, fungsi spline berderajat p adalah sembarang fungsi yang dapat disajikan dalam bentuk :

$$f(x) = \sum_{l=0}^p \alpha_l x^l + \sum_{j=1}^K \beta_j (x - k_j)_+^p$$

Dengan fungsi terpotong sebagai berikut :

$$(x - k_j)^p = \begin{cases} (x - k)^p, & x \geq k_j \\ 0, & x < k_j \end{cases}$$

Dimana $\alpha_l, l = 0, 1, \dots, p$ dan $j = 1, 2, \dots, K$ merupakan parameter dengan K titik knot k_1, k_2, \dots, k_K .

α_l : Koefisien pengganda nilai x^l terhadap variabel respon.

x^l : Variabel prediktor ke- l .

β_j : Koefisien pada variabel x^l truncated knot ke- j pada spline berorde K .

k_j : Knot ke- j pada variabel x^l

K : Banyaknya knot dalam variabel prediktor ke- l .

F. Regresi Spline

Regresi spline adalah regresi dimana kurva regresinya didekati oleh fungsi spline. Fungsi spline adalah fungsi polinomial yang lebih khusus dan memberikan fleksibilitas yang lebih tinggi daripada fungsi polinomial pada umumnya. Fungsi spline yang berhubungan dengan regresi spline terkait adalah fungsi spline linear, kuadrat, dan kubik.

Secara umum, model regresi nonparametrik spline dapat dituliskan sebagai berikut :

$$y_i = \sum_{l=0}^p \alpha_l x^l + \sum_{j=1}^K \beta_j (x - k_j)_+^p + \varepsilon_i,$$

Dimana, $l = 0, 1, \dots, p$ adalah derajat polinomial dan K adalah banyak titik knot pada fungsi *truncated*, serta ε_i merupakan error acak independen dengan mean nol dan varians σ^2 .

G. Regresi Nonparametrik Birespon Spline

Dalam regresi nonparametrik spline jika terdapat satu variabel respon dan satu variabel prediktor maka diperoleh regresi nonparametrik spline univariabel. Jika dalam analisis regresi terdapat satu variabel respon dengan variabel prediktor

lebih dari satu, maka regresi tersebut dinamakan regresi nonparametrik spline multivariabel. Sedangkan regresi birespon didefinisikan sebagai salah satu model regresi yang memiliki variabel respon yang lebih dari satu buah dan diantara variabel-variabel tersebut terdapat korelasi atau hubungan yang kuat, baik secara logika maupun matematis (Simillia, 2007).

Jika regresi birespon memiliki bentuk kurva regresi yang tidak diketahui, maka pendekatan yang digunakan adalah nonparametrik sehingga dikatakan regresi nonparametrik birespon. Model regresi nonparametrik birespon spline dapat dituliskan sebagai berikut :

$$\begin{aligned} y_{1j} &= f_1(x_{1j}) + g_1(x_{2j}) + \varepsilon_{1j} \\ y_{2j} &= f_2(x_{1j}) + g_2(x_{2j}) + \varepsilon_{2j} \end{aligned}$$

Bentuk kurva regresi $f_1(x_{1j})$, $g_1(x_{2j})$, $f_2(x_{1j})$, dan $g_2(x_{2j})$ diasumsikan tidak diketahui. Error random ε_{1j} dan ε_{2j} saling berkorelasi. Dimana error random $\varepsilon_{1j}, j = 1, 2, \dots, n$ saling independen dengan mean nol dan variansi σ_1^2 , dan $\varepsilon_{2j}, j = 1, 2, \dots, n$ saling independen dengan mean nol dan variansi σ_2^2 .

Bentuk umum fungsi spline truncated $f(x)$ derajat p dengan K titik knot k_1, k_2, \dots, k_K dan dua pediktor diberikan sebagai berikut :

$$f(x) = \sum_{i=1}^p \alpha_i x_1^i + \sum_{j=1}^K \beta_j (x_1 - k_j)_+^p + \sum_{i=1}^p \alpha_i^* x_2^i + \sum_{j=1}^K \beta_j^* (x_2 - k_j)_+^p$$

Secara umum model regresi nonparametrik spline dengan derajat p dan K titik knot dapat ditulis sebagai berikut :

$$\begin{aligned} y_i &= \sum_{t=1}^p \alpha_t x_1^t + \sum_{j=1}^K \beta_j (x_1 - k_j)_+^p + \sum_{t=1}^p \alpha_t^* x_2^t \\ &+ \sum_{j=1}^K \beta_j^* (x_2 - k_j)_+^p + \varepsilon_i, \quad i = 1, 2, \dots, n \end{aligned}$$

Berdasarkan bentuk umum model birespon Spline linear dengan satu titik knot dari persamaan adalah sebagai berikut :

$$\begin{aligned} y_{1j} &= \alpha_1 x_{1j} + \beta_1 (x_{1j} - k_1) + \alpha_1^* x_{2j} + \beta_1^* (x_{2j} - k_2) + \varepsilon_{1j} \\ y_{2j} &= \gamma_1 x_{1j} + \delta_1 (x_{1j} - \lambda_1) + \gamma_1^* x_{2j} + \delta_1^* (x_{2j} - \lambda_2) + \varepsilon_{2j} \end{aligned}$$

Bentuk umum model birespon Spline linear kuadratik dengan satu titik knot dari persamaan adalah sebagai berikut :

$$\begin{aligned} y_{1j} &= \alpha_1 x_{1j} + \alpha_2 x_{1j}^2 + \beta_1 (x_{1j} - k_1)_+^2 + \alpha_1^* x_{2j} + \alpha_2^* x_{2j}^2 \\ &+ \beta_1^* (x_{2j} - k_1^*)_+^2 + \varepsilon_{1j} . \\ y_{2j} &= \gamma_1 x_{1j} + \gamma_2 x_{1j}^2 + \delta_1 (x_{1j} - \lambda)_+^2 + \gamma_1^* x_{2j} + \gamma_2^* x_{2j}^2 \\ &+ \delta_1^* (x_{2j} - \lambda_1^*)_+^2 + \varepsilon_{2j} . \end{aligned}$$

Berikut model regresi nonparametrik birespon spline jika disajikan dalam bentuk matriks:

$$\begin{pmatrix} y_1 \\ y_1 \\ \vdots \\ y_{1n} \\ y_2 \\ y_2 \\ \vdots \\ y_{2n} \end{pmatrix} = \begin{pmatrix} f_1(x_1) + f_1(x_2) \\ f_1(x_1) + f_1(x_2) \\ \vdots \\ f_1(x_{1n}) + f_1(x_{2n}) \\ f_2(x_1) + f_2(x_2) \\ f_2(x_1) + f_2(x_2) \\ \vdots \\ f_2(x_{1n}) + f_2(x_{2n}) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_2 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{2n} \end{pmatrix}$$

H. Taksiran Regresi Nonparametrik Birespon Spline

Dengan menguraikan fungsi f dan memisahkan antara parameter dengan variabel, maka persamaan tersebut dalam bentuk matrik dapat ditulis sebagai berikut :

$$y_{\sim} = x\beta_{\sim} + \varepsilon_{\sim}$$

Dimana:

$$y_{\sim} = \begin{pmatrix} y_1 \\ y_1 \\ \vdots \\ y_{1n} \\ y_2 \\ y_2 \\ \vdots \\ y_{2n} \end{pmatrix}; x = \begin{pmatrix} A|O \\ -|- \\ O|B \end{pmatrix} = \beta_{\sim} = \begin{pmatrix} \beta_{1\sim} \\ \beta_{2\sim} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \alpha_1^* \\ \alpha_2^* \\ \beta_1^* \\ \beta_2^* \\ \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \\ \gamma_1^* \\ \gamma_2^* \\ \delta_1^* \\ \delta_2^* \end{pmatrix}; \varepsilon_{\sim} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_2 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{2n} \end{pmatrix}$$

Maka untuk memperoleh penaksir β_{\sim} , dilakukan optimasi WLS yaitu dengan menyelesaikan persamaan sebagai berikut :

$$\beta_{\sim}^m \{ \epsilon_{\sim}' W^{-1} \epsilon_{\sim} \} = \beta_{\sim}^m \{ (y_{\sim} - x_{\sim})' W^{-1} (y_{\sim} - x_{\sim}) \}$$

Untuk menyelesaikan optimasi pada persamaan maka dilakukan derivatif parsial. Dengan memisalkan fungsi $Q(\beta_{\sim}) = (y_{\sim} - x_{\sim})' W^{-1} (y_{\sim} - x_{\sim})$, maka

$$\begin{aligned} Q(\beta_{\sim}) &= (y_{\sim} - x_{\sim})' W^{-1} (y_{\sim} - x_{\sim}) \\ &= (y_{\sim}' - \beta_{\sim}' x_{\sim}')' W^{-1} (y_{\sim} - x_{\sim}) \\ &= (y_{\sim}' - \beta_{\sim}' x_{\sim}')' (W^{-1} y_{\sim} - W^{-1} x_{\sim}) \\ &= y_{\sim}' W^{-1} y_{\sim} - \beta_{\sim}' x_{\sim}' W^{-1} y_{\sim} - y_{\sim}' W^{-1} x_{\sim} \\ &\quad + \beta_{\sim}' x_{\sim}' W^{-1} x_{\sim} \\ &= y_{\sim}' W^{-1} y_{\sim} - 2\beta_{\sim}' x_{\sim}' W^{-1} y_{\sim} \\ &\quad + \beta_{\sim}' x_{\sim}' W^{-1} x_{\sim} \end{aligned}$$

Proses selanjutnya adalah menurunkan persamaan terhadap β_{\sim} dan dihasilkan :

$$\begin{aligned} \frac{\partial Q(\beta_{\sim})}{\partial \beta_{\sim}} &= \frac{\partial (y_{\sim}' W^{-1} y_{\sim} - 2\beta_{\sim}' x_{\sim}' W^{-1} y_{\sim} + \beta_{\sim}' x_{\sim}' W^{-1} x_{\sim})}{\partial \beta_{\sim}} \\ &= -2x_{\sim}' W^{-1} y_{\sim} + 2x_{\sim}' W^{-1} x_{\sim} \end{aligned} \quad (3.7)$$

Setelah diturunkan terhadap β_{\sim} hasil (3.7) disamakan dengan nol.

$$\begin{aligned} -2x_{\sim}' W^{-1} y_{\sim} + 2x_{\sim}' W^{-1} x_{\sim} &= 0 \\ x_{\sim}' W^{-1} x_{\sim} \hat{\beta}_{\sim} &= x_{\sim}' W^{-1} y_{\sim} \end{aligned}$$

Berdasarkan persamaan (3.8), didapatkan penaksir $\hat{\beta}_{\sim}$ sebagai berikut :

$$\hat{\beta}_{\sim} = (x_{\sim}' W^{-1} x_{\sim})^{-1} x_{\sim}' W^{-1} y_{\sim}$$

Sehingga bentuk taksiran model spline dalam regresi nonparametrik birespon spline menjadi sebagai berikut :

$$\hat{y}_{\sim} = (x_{\sim}' W^{-1} x_{\sim})^{-1} x_{\sim}' W^{-1} y_{\sim}$$

Jika matriks $H(k_{\sim}) = x_{\sim} (x_{\sim}' W^{-1} x_{\sim})^{-1} x_{\sim}' W^{-1}$, maka diperoleh $\hat{y}_{\sim} = H(k_{\sim}) y_{\sim}$. Matrik $H(k_{\sim})$ merupakan fungsi dari titik knot, sedangkan $k_{\sim} = (k_1, k_2, \dots, k_K)'$ adalah titik-titik knot.

I. Pemilihan Titik Knot Optimal dan Model Terbaik

Model spline yang terbaik adalah model dengan titik knot yang optimal, dimana model ini merupakan model yang paling sesuai dengan data. Menurut (Wahba, 1990) dan (Wang, 1998), salah satu metode yang paling banyak dipakai dan disukai karena kelebihan yang dimilikinya adalah *Generalized Cross Validation* (GCV). Dibandingkan dengan metode lain, misal *Cross Validation* (CV), metode GCV mempunyai sifat optimal

asimtotik (Wahba, 1990). Titik knot yang optimal diperoleh nilai GCV yang paling minimum. Model terbaik dipilih berdasarkan nilai GCV yang paling minimum. Kriteria sederhana yang digunakan sebagai ukuran kinerja atas penaksir yang baik adalah *Mean Square Error* (MSE).

$$M(K) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_K(x_i))^2$$

Kriteria lain yang dapat digunakan sebagai ukuran kinerja atas penaksir yang baik adalah *Generalized Cross Validation*. Seperti yang dituliskan pada persamaan berikut :

$$G(K) = \frac{1}{n} \sum_{i=1}^n \left[\frac{y_i - \hat{f}_K(x_i)}{\{n^{-1}(t_i(I) - t_i(H(K)))\}^2} \right]^2$$

$$= \frac{n^{-1} \sum_{j=1}^n (y_j \hat{f}_K(x_j))^2}{\{n^{-1}(t_i(I) - t_i(H(K)))\}^2}$$

dengan $H(K)$ merupakan matriks hat, yaitu $H = H(K) = X(X'W^{-1}X)^{-1}X'W^{-1}$

3. HASIL DAN PEMBAHASAN

A. Sumber Data

Data dalam skripsi ini merupakan data sekunder yaitu data Angka Melek Huruf (Y1), Rata-Rata Lama Sekolah (Y2) dan Rata-Rata Jumlah Anggota Keluarga (X1) yang diambil dari BPS. Dan data Pengeluaran Anggaran Pemerintah Daerah Untuk Pendidikan (X2) dari Departemen Keuangan Republik Indonesia. Data yang diperoleh adalah data tahun 2013 di 26 Kabupaten dan Kota di Provinsi Jawa Barat.

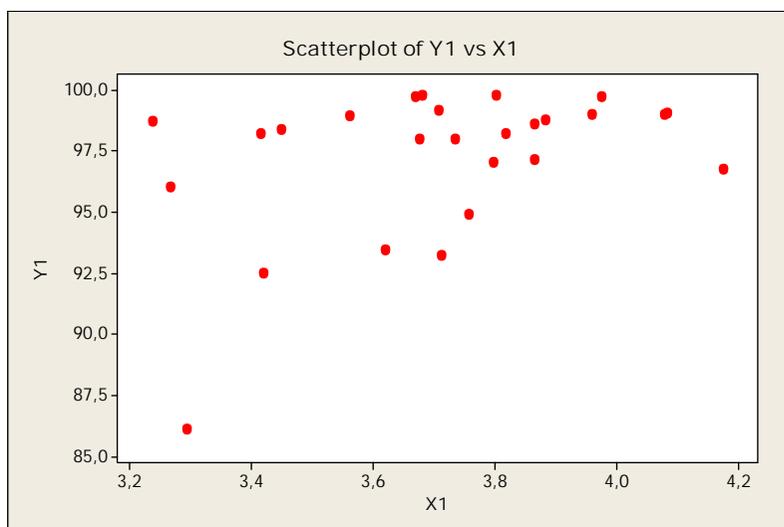
B. Analisis Deskriptif

Sebelum dilakukan pemilihan model terbaik menggunakan regresi birespon spline perlu dilakukan analisis deskriptif. Berikut analisis untuk data Angka Melek Huruf (Y1) , Rata-Rata Lama Sekolah (Y2), Rata-Rata Anggota Rumah Tangga (X1) dan Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan (X2) .

Tabel 1. Analisis Deskriptif

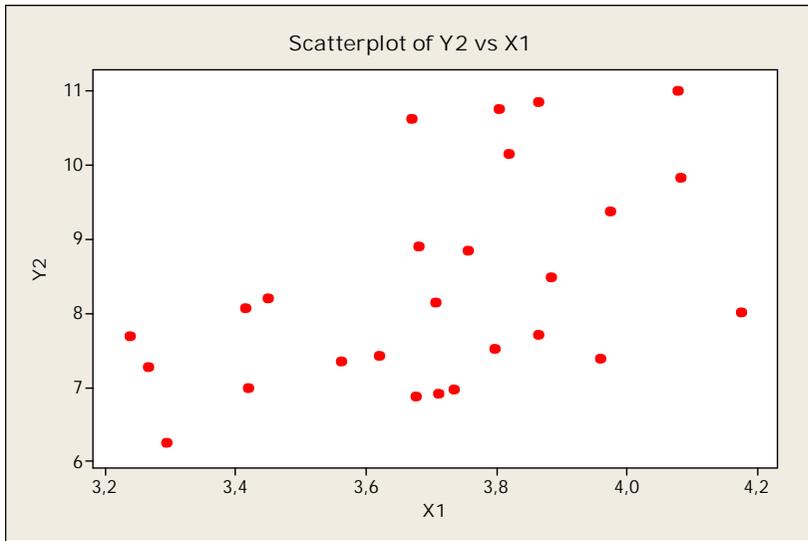
Variabel	Rata-rata	Standar Deviasi	Minimum	Maksimum
Y1	97,26	3,07	86,11	99,82
Y2	8,37	1,41	6,25	11
X1	3,71	0,25	3,24	4,18
X2	465933,58	276280,25	52120	1245963

Dalam penelitian ini terdapat korelasi antara Angka Melek Huruf (Y1) dan Rata-Rata Lama Sekolah (Y2) yaitu sebesar 0,561. Adapun pola data dalam penelitian ini digambarkan dengan menggunakan *scatter plot* yaitu sebagai berikut :



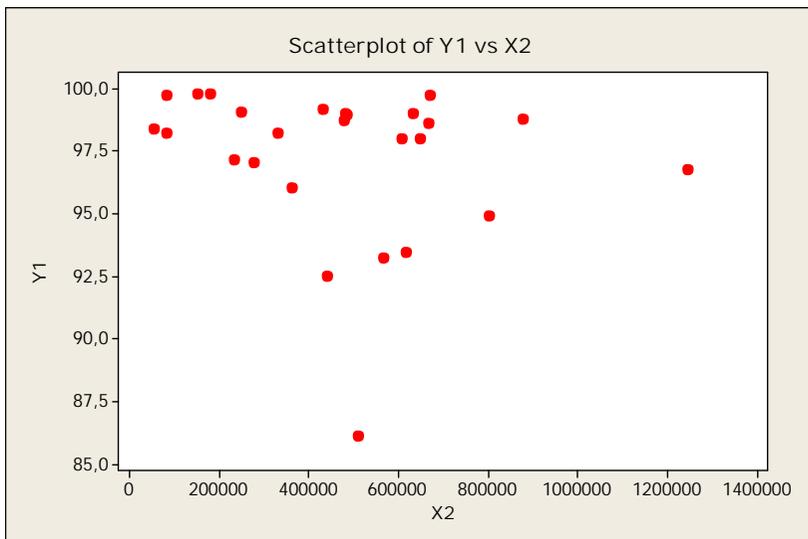
Gambar 1. *Scatter Plot* Angka Melek Huruf (Y1) dan Rata-Rata Anggota Rumah Tangga (X1)

Berdasarkan Gambar 1 yaitu *scatter plot* Angka Melek Huruf (Y1) dan Rata-Rata Anggota Rumah Tangga (X1) terlihat pola datanya tidak mengikuti pola data tertentu.



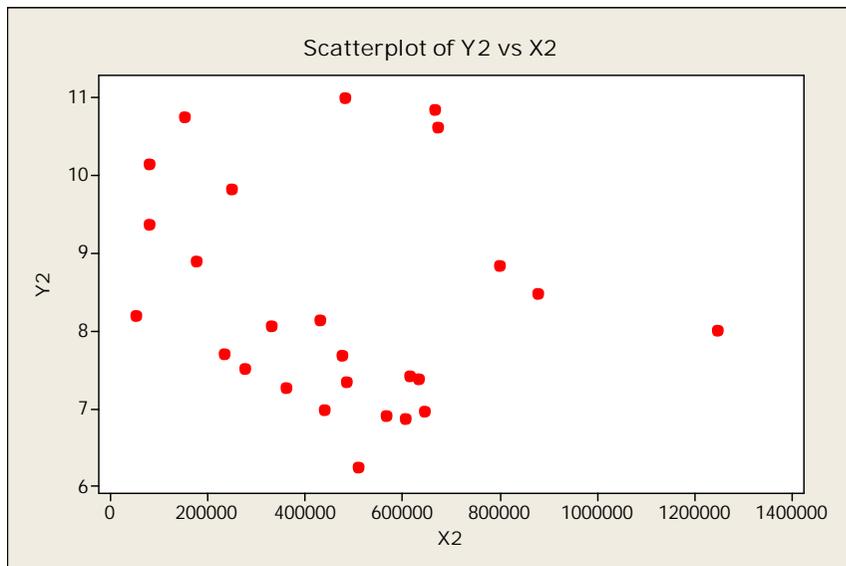
Gambar 2. *Scatter Plot* Rata-Rata Lama Sekolah (Y2) dan Rata-Rata Anggota Rumah Tangga (X1)

Berdasarkan Gambar 2 yaitu *scatter plot* Rata-Rata Lama Sekolah (Y2) dan Rata-Rata Anggota Rumah Tangga (X1) terlihat pola datanya tidak mengikuti pola data tertentu.



Gambar 3. *Scatter Plot* Angka Melek Huruf (Y1) dan Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan (X2)

Berdasarkan Gambar 3 yaitu *Scatter Plot* Angka Melek Huruf (Y1) dan Pengeluaran Pemerintah Daerah pada Bidang Pendidikan (X2) terlihat pola datanya tidak mengikuti pola data tertentu.



Gambar 4 *Scatter Plot* Rata-Rata Lama Sekolah (Y2) dan Pemerintah Daerah Pada Bidang Pendidikan (X2)

Berdasarkan Gambar 4.4 yaitu *scatter plot* Rata-Rata Lama Sekolah (Y2) dan Pemerintah Daerah Pada Bidang Pendidikan (X2) terlihat pola datanya tidak mengikuti pola data tertentu.

Berdasarkan *scatter plot* di atas pola data tidak menunjukkan adanya pola tertentu, sehingga dalam penelitian ini maka digunakan pendekatan regresi nonparametrik birespon spline.

C. Model Spline Linear Birespon Dengan Satu Titik Knot

Berdasarkan bentuk umum model birespon Spline linear dengan satu titik knot adalah sebagai berikut :

$$y_{1j} = \alpha_1 x_{1j} + \beta_1 (x_{1j} - k_1) + \alpha_1^* x_{2j} + \beta_1^* (x_{2j} - k_2) + \varepsilon_{1j}$$

$$y_{2j} = \gamma_1 x_{1j} + \delta_1 (x_{1j} - \lambda_1) + \gamma_1^* x_{2j} + \delta_1^* (x_{2j} - \lambda_2) + \varepsilon_{2j}$$

Pemilihan model birespon spline terbaik diperoleh berdasarkan nilai GCV (*Generalized Cross Validation*) yang minimum.

Berdasarkan Tabel maka estimasi model spline linear birespon dengan satu titik knot dapat ditulis kedalam bentuk persamaan sebagai berikut:

$$\hat{y}_1 = 1,20436x_1 - 8,25187(x_1 - 3,6) - 2,46275x_2 + 1,06604(x_2 - 4).$$

$$\hat{y}_2 = -1,33375x_1 - 2,79147(x_1 - 4) - 8,84373x_2 + 1,09253(x_2 - 21).$$

D. Model Spline Kuadratik Birespon Dengan Satu Titik Knot

Bentuk umum model birespon Spline kuadratik dengan satu titik knot adalah sebagai berikut :

$$y_{1j} = \alpha_1 x_{1j} + \alpha_2 x_{1j}^2 + \beta_1 (x_{1j} - k_1)_+^2 + \alpha_1^* x_{2j} + \alpha_2^* x_{2j}^2 + \beta_1^* (x_{2j} - k_1^*)_+^2 + \varepsilon_{1j}.$$

$$y_{2j} = \gamma_1 x_{1j} + \gamma_2 x_{1j}^2 + \delta_1 (x_{1j} - \lambda)_+^2 + \gamma_1^* x_{2j} + \gamma_2^* x_{2j}^2 + \delta_1^* (x_{2j} - \lambda_1^*)_+^2 + \varepsilon_{2j}.$$

Seperti pada model spline linear, pemilihan model birespon spline terbaik diperoleh berdasarkan nilai GCV yang minimum.

Berdasarkan Tabel maka taksiran model spline linear birespon dengan satu titik knot dapat ditulis kedalam bentuk persamaan sebagai berikut:

$$\hat{y}_1 = -4.97767x_1 - 6.54222x_1^2 - 8.36634(x_1 - 3,5)_+^2 + 9.37356x_2 - 1.98746x_2^2 - 2.08476(x_2 - 8)_+^2$$

$$\hat{y}_2 = 4.93650x_1 - 1.08746x_1^2 - 2.04899(x_1 - 4)_+^2 + 5.00847x_2 - 7.03839x_2^2 - 2.33575(x_2 - 34)_+^2$$

E. Model Spline Birespon Optimal

Setelah mendapatkan model spline dalam egresi nonparametrik birespon menggunakan bentuk linear dan kuadratik dengan masing-masing satu titik knot yang berbeda, maka didapatkan model birespon spline yang terbaik dilihat berdasarkan nilai GCV yang minimum.

Tabel 2 Nilai GCV pada masing-masing Model Spline Birespon

Model Spline	Nilai GCV
Linear Satu Knot	0,026876*
Kuadratik Satu Knot	0,034975

Berdasarkan Tabel 2 dapat diketahui bahwa nilai GCV yang paling minimum sebesar 0,026876. Sehingga model terbaik yang menjelaskan data Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat adalah model Spline Linear dengan satu titik knot mempunyai bentuk model sebagai berikut:

$$\begin{aligned}\hat{y}_1 &= 1,20436x_1 - 8,25187(x_1 - 3,6) - 2,46275x_2 \\ &\quad + 1,06604(x_2 - 4) \\ \hat{y}_2 &= -1,33375x_1 - 2,79147(x_1 - 4) - 8,84373x_2 \\ &\quad + 1,09253(x_2 - 21)\end{aligned}$$

F. Interpretasi Model

Untuk memudahkan interpretasi model regresi nonparametrik birespon spline data Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat, maka dibuat model parsialnya dari model optimal yang telah dibentuk.

Dengan asumsi X_2 konstan, maka pengaruh dari Rata-Rata Banyaknya Anggota Rumah Tangga (X_1) adalah sebagai berikut.

$$\begin{aligned}\hat{y}_1 &= 1,20436x_1 - 8,25187(x_1 - 3,6) \\ \hat{y}_1 &= \begin{cases} 1,20436x_1 & ; x_1 < 3,6 \\ -7,047339x_1 + 4,335696 & ; x_1 \geq 3,6 \end{cases}\end{aligned}$$

Berdasarkan model yang terbentuk di atas dapat menjelaskan bahwa untuk Rata-Rata Banyaknya Anggota Rumah Tangga kurang dari 3,6 artinya adalah jika Rata-Rata Banyaknya Anggota Rumah Tangga naik sebesar satu persen maka jumlah Angka Melek Huruf di Jawa Barat naik sebesar 1,20436. Sedangkan untuk Rata-Rata Banyaknya Anggota Rumah Tangga lebih besar atau sama dengan 3,6 artinya adalah jika Rata-Rata Banyaknya Anggota Rumah Tangga meningkat sebesar satu persen, maka Angka Melek Huruf di Jawa Barat menurun sebesar 7,047339.

Untuk respon kedua yakni Rata-Rata Lama Sekolah

$$\begin{aligned}\hat{y}_2 &= -1,33375x_1 - 2,79147(x_1 - 4) \\ \hat{y}_2 &= \begin{cases} -1,33375x_1 & ; x_1 < 4 \\ -4,12522x_1 + 5,335 & ; x_1 \geq 4 \end{cases}\end{aligned}$$

Berdasarkan model yang terbentuk di atas dapat dijelaskan bahwa untuk Rata-Rata Banyaknya Anggota Rumah Tangga kurang dari 4 artinya adalah jika Rata-Rata Banyaknya Anggota Rumah Tangga naik sebesar satu persen maka Rata-Rata Lama Sekolah di Jawa Barat menurun sebesar 1,33375. Sedangkan untuk Rata-Rata Banyaknya Anggota Rumah Tangga lebih besar atau sama dengan 4 artinya adalah jika Rata-Rata Banyaknya Anggota Rumah Tangga meningkat sebesar satu persen, maka Rata-Rata Lama Sekolah di Jawa Barat menurun sebesar 4,12522.

Dengan asumsi X_1 konstan, maka pengaruh dari Pengeluaran Pemerintah Daerah pada Bidang Pendidikan (x_2) adalah sebagai berikut.

$$\begin{aligned}\hat{y}_1 &= -2,46275x_2 + 1,06604(x_2 - 4) \\ \hat{y}_1 &= \begin{cases} 2,46275x_2 & ; x_2 < 4 \\ -1,39671x_2 - 4,26416 & ; x_2 \geq 4 \end{cases}\end{aligned}$$

Berdasarkan model yang terbentuk di atas dapat dijelaskan bahwa untuk Pengeluaran Pemerintah Daerah pada Bidang Pendidikan kurang dari 4 artinya adalah jika Pengeluaran Pemerintah Daerah pada Bidang Pendidikan naik sebesar satu persen maka jumlah Angka Melek Huruf di Jawa Barat naik sebesar 2,46275. Sedangkan untuk Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan lebih besar atau sama dengan 4 artinya adalah jika Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan meningkat sebesar satu persen, maka Angka Melek Huruf di Jawa Barat menurun sebesar 1,39671

Untuk respon kedua yakni Rata-Rata Lama Sekolah

$$\hat{y}_2 = -8,84373x_2 + 1,09253(x_2 - 21)$$

$$\hat{y}_2 \begin{cases} -8,84373x_2 & ; x_2 < 21 \\ -7,7512x_1 + 22,94313 & ; x_2 \geq 21 \end{cases}$$

Berdasarkan model yang terbentuk di atas dapat dijelaskan bahwa untuk Pengeluaran Pemerintah Daerah pada Bidang Pendidikan kurang dari 21 artinya adalah jika Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan naik sebesar satu persen maka Rata-Rata Lama Sekolah di Jawa Barat menurun sebesar 8,84373. Sedangkan untuk Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan lebih besar atau sama dengan 21 artinya adalah jika Pengeluaran Pemerintah Daerah Pada Bidang Pendidikan meningkat sebesar satu persen, maka Rata-Rata Lama Sekolah di Jawa Barat menurun sebesar 7,7512.

4. KESIMPULAN DAN SARAN

1. Model Regresi Nonparametrik Birespon Spline sebagai berikut:

$$y_{1j} = \alpha_1 x_{1j} + \beta_1 (x_{1j} - k_1) + \alpha_1^* x_{2j} + \beta_1^* (x_{2j} - k_2) + \varepsilon_{1j}$$

$$y_{2j} = \gamma_1 x_{1j} + \delta_1 (x_{1j} - \lambda_1) + \gamma_1^* x_{2j} + \delta_1^* (x_{2j} - \lambda_2) + \varepsilon_{2j}$$

Taksiran model Spline dalam Regresi Nonparametrik Birespon adalah $\hat{y}_\sim = H(k_\sim)y_\sim$ dengan matrik $H(k_\sim) = x(x'W^{-1}x)^{-1}x'W^{-1}$

2. Model Spline Birespon terbaik dengan satu titik knot yang menjelaskan data Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat mempunyai bentuk taksirannya sebagai berikut:

$$\hat{y}_1 = 1,20436x_1 - 8,25187(x_1 - 3,6) - 2,46275x_2 + 1,06604(x_2 - 4)$$

$$\hat{y}_2 = -1,33375x_1 - 2,79147(x_1 - 4) - 8,84373x_2 + 1,09253(x_2 - 21)$$

Saran yang diberikan diharapkan untuk peneliti berikutnya agar menambahkan model spline birespon selain linear dan kuadrat dengan kubik dan dengan menambah jumlah titik knot lebih dari satu. Untuk Pemerintah diharapkan agar lebih memperhatikan tingkat pendidikan khususnya di Jawa Barat.

5. DAFTAR PUSTAKA

- [1] BPS. (2016). *Indeks Pembangunan Manusia*. Dipetik Februari 24, 2016, dari BPS: <http://bps.go.id/Subjek/view/id/26#subjekViewTab1|accordion-daftar-subjek1>.
- [2] Budiantara, I. N. (2009). *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah pemodelan Statistika Masa Kini dan Masa Mendatang*.
- [3] Budiantara, I. N. (2009). *Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter*.
- [4] Draper, N. d. (1998). Dalam *Applied Regression Analysis Third Edition*. Canada: A Wiley Interscience Publication.
- [5] Eubank. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- [6] Fox, J. (2005). Introduction to Nonparametric Regression. Dalam J. Fox, *Introduction to Nonparametric Regression*. Canada: University of Oxford.
- [7] Indonesia, K. K. (n.d.). *Data Publikasi*. Retrieved Februari 20, 2016, from Kementrian Keuangan Republik Indonesia: www.kemenkeu.go.id

- [8] Oktaviana, D. (2011). Regresi Spline Birespon untuk Memodelkan Kadar Gula Darah Penderita Diabetes Melitus. *Regresi Spline Birespon untuk Memodelkan Kadar Gula Darah Penderita Diabetes Melitus*.
- [9] Simillia, T. (2007). *Input Selection and Shrinkage in Multiresponse Linear Regression*.
- [10] Setyawan, N. A. (2009). Nonparametric Biresponse Spline Regression Approach on Modeling Determinants of Education Outcome in Papua Island. *Nonparametric Biresponse Spline Regression Approach on Modeling Determinants of Education Outcome in Papua Island*.
- [11] Stone, J. Z. (2002). Extended Linear Modeling With Splines. *Extended Linear Modeling With Splines*.
- [12] Tripena, A. (2011). Penentuan Model Regresi Spline Terbaik. *Penentuan Model Regresi Spline Terbaik*.
- [13] Wahba. (1990). Dalam Wahba, *Spline Models for Observational Data*. SIAM Pennsylvania.
- [14] Wang. (1998). Dalam Wang, *Spline Smoothing Models With Correlated Errors*. Journal of the American Statistical Association.
- [15] Wulandari, I. D. (2011). Analisis Faktor-Faktor yang Mempengaruhi Presentase Penduduk Miskin dan Pengeluaran Perkapita Makanan di Jawa Timur Menggunakan Regresi Nonparametrik Birespon Spline. *Analisis Faktor-Faktor yang Mempengaruhi Presentase Penduduk Miskin dan Pengeluaran Perkapita Makanan di Jawa Timur Menggunakan Regresi Nonparametrik Birespon Spline*.
- [16] Yuyun Yuniartika, D. K. (2013). Penentuan Generalized Cross Validation (GCV) Sebagai Kriteria dalam Pemilihan Model Regresi B-SPLINE Terbaik.