

PEMODELAN RATA-RATA LAMA SEKOLAH MENGGUNAKAN PENDEKATAN REGRESI NONPARAMETRIK *SPLINE* *MULTIVARIABLE*

Ni Luh Gede Sinta Aryati[§], I Komang Gde Sukarsa², I Gusti Ayu Made Srinadi³

¹Program Studi Matematika, Fakultas MIPA - Universitas Udayana [Email : sintaaryati@gmail.com]

²Program Studi Matematika, Fakultas MIPA - Universitas Udayana [Email : gedesukarsa@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA - Universitas Udayana [Email : srinadi@unud.ac.id]

[§] Corresponding Author

ABSTRACT

Mean years school (MYS) is one of the indicators used in calculating the human development index (HDI). The value of MYS Indonesia in 2019 is 8,75 which is still low. Therefore it still needs to be improved. In this research, MYS modeling will be carried out using six factors that are thought to influence MYS. This research uses multivariable spline nonparametric regression to modeling MYS Indonesia in 2019. The best model is selected based on the minimum value of Generalized Cross Validation (GCV). Based on this research, the best model obtained is a linear orde (orde 2) spline model with four knots. The model has R^2 value of 99,91%.

Keywords: Mean Years School (MYS), Nonparametric, Spline, Knot, GCV.

1. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi mengakibatkan persaingan antarnegara menjadi semakin ketat. Pendidikan merupakan salah satu indikator yang perlu diperhatikan untuk meningkatkan kualitas sumber daya manusia agar dapat bersaing di era revolusi 4.0 ini. Kualitas sumber daya manusia dapat dilihat melalui indeks pembangunan manusia (IPM). Salah satu indikator IPM dalam bidang pendidikan yaitu rata-rata lama sekolah (RLS). RLS menunjukkan banyak tahun yang telah diselesaikan untuk menempuh pendidikan formal (BPS, 2020). Semakin besar RLS maka semakin banyak masyarakat yang dapat menyelesaikan pendidikan di tingkat yang lebih tinggi sehingga diharapkan dapat meningkatkan kualitas pendidikan masyarakat. Pada penelitian ini dilakukan pemodelan RLS masing-masing provinsi di Indonesia pada tahun 2019 menggunakan beberapa faktor yang diduga berpengaruh terhadap RLS. Terdapat enam faktor yang diduga dapat berpengaruh terhadap RLS yaitu persentase penduduk miskin, proporsi disabilitas anak, angka melek huruf, angka putus sekolah jenjang SD, angka putus sekolah jenjang SMP dan angka putus sekolah jenjang SMA.

Adapun metode statistika yang digunakan adalah regresi. Analisis regresi dapat berupa regresi parametrik, semiparametrik atau nonparametrik. Regresi parametrik mensyaratkan adanya asumsi mengenai bentuk kurva regresi. Akan tetapi terkadang bentuk kurva regresi sulit untuk diasumsikan. Oleh karena itu akan lebih baik jika menggunakan analisis regresi nonparametrik karena tidak mensyaratkan asumsi terhadap bentuk kurva regresi. Salah satu regresi nonparametrik yaitu regresi nonparametrik *spline*. *Spline* memberikan fleksibilitas yang lebih baik terhadap karakteristik suatu fungsi dengan membiarkan data mencari bentuk fungsi hubungannya sendiri. Selain itu *spline* juga mampu menjelaskan pola perilaku fungsi dalam sub-interval tertentu dan dapat digunakan untuk mengatasi atau mengurangi pola data yang mengalami peningkatan tajam dengan bantuan titik knot (Griggs, 2013). *Spline* bergantung pada penentuan titik *knot* optimal menggunakan nilai *generalized cross validation* (GCV) yang paling kecil.

Penelitian terdahulu yang menggunakan analisis regresi nonparametrik *spline* dilakukan oleh Prahutama et al. (2014) yang memodelkan nilai inflasi berdasarkan harga-harga pangan

menggunakan *spline multivariable*, didapat model dengan nilai koefisien determinasi (R^2) tertinggi yaitu sebesar 0,9394 pada orde 2 dengan 1 titik knot. Penelitian lainnya dilakukan oleh Astiti et al. (2016) yang memodelkan kemiskinan di Indonesia menggunakan analisis *spline multivariable*. Kesimpulan yang diperoleh dari penelitian tersebut adalah model terbaik model regresi *spline* kubik dengan lima titik knot dan nilai R^2 sebesar 0,9999. Beberapa penelitian tersebut menunjukkan bahwa regresi nonparametrik *spline multivariable* memberikan model yang baik dan dengan koefisien determinasi yang mendekati 1. Adapun tujuan dari penelitian ini yaitu mengetahui model yang tepat untuk memodelkan rata-rata lama sekolah di Indonesia menggunakan pendekatan regresi nonparametrik *spline multivariable*.

Regresi secara umum dapat dituliskan sebagai berikut (Kurtner & Nachtsheim, 2004) :

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

dimana Y_i adalah nilai variabel respons dalam amatan ke- i , X_i adalah peubah bebas atau variabel prediktor dalam amatan ke- i , $f(X_i)$ adalah regresi yang telah diketahui bentuknya, ε_i adalah suku galat atau sisaan yang diasumsikan independen dan bersifat acak dengan nilai tengah nol dan variansi σ^2 dalam amatan ke- i dan n adalah banyaknya amatan.

Regresi nonparametrik dapat dirumuskan sebagai berikut (Eubank, 1999) :

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

dengan y_i adalah variabel respons dalam amatan ke- i , $f(t_i)$ adalah fungsi kurva regresi yang tidak diketahui bentuknya, t_i adalah variabel prediktor, ε_i adalah galat yang diasumsikan berdistribusi $N(0, \sigma^2)$.

Regresi nonparametrik *spline* dapat dituliskan sebagai berikut (Eubank, 1999):

$$y = \sum_{j=0}^{p-1} \beta_j x_i^j + \sum_{l=1}^r \beta_{(p+l)} (x_i - k_l)_+^p \quad (3)$$

Estimasi regresi nonparametrik *spline* diperoleh dengan menggunakan metode *maximum likelihood estimation* (MLE). Persamaan (2) respons Y_i diasumsikan berdistribusi normal dengan nilai tengah $f(t_i)$ dan variansi σ^2 , sehingga fungsi densitas peluang Y_i dengan $f(t) > 0, \sigma^2 > 0$. Estimasi

parameter *spline* dapat dituliskan sebagai berikut :

$$\hat{Y} = x(x'x)^{-1}x'Y. \quad (4)$$

Terdapat beberapa metode yang digunakan untuk memilih titik *knot* optimal, salah satunya yaitu dengan metode GCV. Titik *knot* optimal dipilih berdasarkan nilai GCV. GCV dapat dirumuskan sebagai berikut (Eubank, 1999):

$$GCV(K) = \frac{MSE(K)}{(n^{-1}tr[I - A(K)])^2} \quad (5)$$

dengan

$$MSE(K) = n^{-1} \sum_{i=1}^n (Y_i - \hat{f}(x_i))^2. \quad (6)$$

Salah satu kriteria yang digunakan untuk menentukan model terbaik yaitu koefisien determinasi (R^2) dirumuskan sebagai berikut :

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2}. \quad (7)$$

2. METODE PENELITIAN

2.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh melalui *website* Badan Pusat Statistik. Data tersebut yaitu data rata-rata lama sekolah masing-masing provinsi di Indonesia tahun 2019 dengan banyaknya observasi yaitu 34 provinsi.

2.2 Variabel Penelitian

Terdapat dua jenis variabel yang digunakan dalam penelitian ini yaitu variabel respon (Y) dan variabel prediktor (X). Variabel respon dalam penelitian ini yaitu rata-rata lama sekolah masing-masing provinsi di Indonesia tahun 2019 (Y). Penelitian ini menggunakan enam variabel prediktor yaitu persentase penduduk miskin (X_1), proporsi disabilitas anak (X_2), angka melek huruf (X_3), angka putus sekolah jenjang SD (X_4), angka putus sekolah jenjang SMP (X_5) dan angka putus sekolah jenjang SMA (X_6).

2.3 Metode Analisis Data

Analisis dalam penelitian ini menggunakan metode analisis regresi nonparametrik *spline multivariable*. Langkah-langkah yang

dilakukan dalam penelitian ini adalah sebagai berikut :

1. Menyajikan statistika deskriptif dari setiap variabel.
2. Memilih titik *knot* optimal berdasarkan GCV yang paling minimum.
3. Memodelkan rata-rata lama sekolah dengan variabel-variabel prediktornya menggunakan regresi nonparametrik *spline* dengan titik *knot* optimal yang diperoleh dari langkah nomor 2.
4. Menghitung nilai koefisien determinasi (R^2) dari model.
5. Membuat interpretasi dari model yang diperoleh dan membuat kesimpulan yang diperoleh berdasarkan model.

3. PEMBAHASAN

3.1 Deskripsi Data

Karakteristik dari masing-masing variabel disajikan dalam statistika deskriptif sebagai berikut.

Tabel 1. Statistika Deskriptif Variabel Penelitian

Variabel	Mean	St. Deviasi	Min	Max
Y	8,962	0,841	6,85	11,11
X_1	10,427	5,437	3,78	26,64
X_2	3,459	1,321	1,40	7,00
X_3	96,340	4,295	7,00	99,80
X_4	0,370	0,369	0,00	2,00
X_5	1,070	0,707	0,18	2,91
X_6	1,760	1,562	0,18	6,10

Rata-rata lama sekolah (Y) tahun 2019 tercatat sebesar 8,75 tahun. Angka tersebut telah mengalami peningkatan sebesar 0,17% dibandingkan tahun sebelumnya namun masih belum dapat memenuhi target yang ditetapkan pemerintah yaitu 9 tahun. DKI Jakarta menempati posisi teratas dengan Y lebih dari 11 tahun yang artinya rata-rata penduduk DKI Jakarta sudah bisa mengenyam pendidikan sampai SMA kelas XII. Di sisi lain, penduduk 15 tahun ke atas di Papua rata-rata bersekolah hanya sampai kelas 6 SD/ sederajat yang artinya rata-rata penduduk Papua hanya mampu mengenyam pendidikan sampai bangku SD.

Variabel X_1 merupakan persentase penduduk miskin. Rata-rata X_1 yaitu sebesar 10,427% menurun 0,44% dibandingkan tahun 2018. X_1 terbesar yaitu Provinsi Papua sedangkan terendah yaitu Provinsi Bali. Nilai standar deviasi sebesar 5,437 menunjukkan bahwa data cukup beragam. Variabel X_2 merupakan proporsi disabilitas anak. Rata-rata X_2 adalah sebesar 3,459% dengan status tidak bersekolah yang lebih besar dibandingkan yang bersekolah. X_2 terbesar yaitu 7% pada Provinsi Sulawesi Tengah sedangkan terendah yaitu 1,4% pada Provinsi Jambi. Variabel X_3 merupakan angka melek huruf. Rata-rata X_3 adalah sebesar 96,34% meningkat sebesar 0,24 persen dibandingkan tahun sebelumnya. X_3 terbesar yaitu Sulawesi Utara sedangkan terendah yaitu Papua. Nilai standar deviasi cukup besar yaitu 4,295 menunjukkan data cukup beragam. Variabel X_4 merupakan angka putus sekolah jenjang SD. Rata-rata X_4 adalah sebesar 0,37. Angka tersebut sudah mampu melampaui target pemerintah yaitu 1%. X_4 terbesar yaitu Papua sedangkan terendah yaitu Kalimantan Utara. Variabel X_5 merupakan angka putus sekolah jenjang SMP. Rata-rata X_5 adalah sebesar 1,07%. Angka tersebut belum mampu mencapai target yang ditetapkan pemerintah yaitu 1%. X_5 terbesar yaitu Sulawesi Barat sedangkan terendah yaitu Papua Barat. Variabel X_6 merupakan angka putus sekolah jenjang SMA. Rata-rata X_6 adalah sebesar 1,76%. Angka tersebut belum mampu mencapai target yang ditetapkan pemerintah yaitu 1%. X_6 terbesar yaitu DKI Jakarta sedangkan X_6 terendah yaitu Nusa Tenggara Barat.

3.2 Pemilihan Titik *Knot*

Pemilihan titik *knot* dalam penelitian ini dilakukan menggunakan metode GCV. Titik *knot* optimal dipilih berdasarkan nilai GCV yang paling minimum. Pada penelitian ini pemilihan titik *knot* dibatasi dari satu titik *knot*, dua titik *knot*, tiga titik *knot*, dan empat titik *knot* dengan *spline* linear (orde 2) dan *spline* kuadratik (orde 3). Hal ini bertujuan untuk mempermudah interpretasi dan menghindari terbentuknya matriks yang singular. Analisis data dilakukan menggunakan *software* R.

Pada Tabel 2. berikut ini disajikan nilai GCV masing-masing titik *knot* beserta orde *knot*.

Tabel 2. Nilai GCV dari Variasi Titik *Knot* dan Orde

Orde	Banyak Titik <i>Knot</i>	GCV
Orde 2	1 Titik <i>Knot</i>	0,20
	2 Titik <i>Knot</i>	0,16
	3 Titik <i>Knot</i>	0,22
	4 Titik <i>Knot</i>	0,07
Orde 3	1 Titik <i>Knot</i>	0,36
	2 Titik <i>Knot</i>	0,36
	3 Titik <i>Knot</i>	0,68
	4 Titik <i>Knot</i>	20924,52

Berdasarkan Tabel 2 dapat diketahui bahwa nilai GCV minimum yaitu 0.07 pada orde 2 dengan empat titik *knot*. Adapaun titik *knot* optimal yang diperoleh yaitu 6,1; 7,88; 12,34; 14,99; 2,5; 3,1; 3,8; 4,8; 93,54; 97,72; 98,75; 99,03; 0,18; 0,32; 0,6; 0,65; 0,32; 0,84; 1,17; dan 1,71. Oleh karena itu, diputuskan bahwa model terbaik yang akan dipilih adalah model *spline* linier (orde 2) dengan 4 titik *knot*.

3.3 Estimasi Parameter Model *Spline*

Estimasi model regresi nonparametrik *spline* orde 2 dengan empat titik *knot* dapat dituliskan sebagai berikut.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (X_1 - K_1)_+^1 + \hat{\beta}_3 (X_1 - K_2)_+^1 + \hat{\beta}_4 (X_1 - K_3)_+^1 + \hat{\beta}_5 (X_1 - K_4)_+^1 + \hat{\beta}_6 X_2 + \hat{\beta}_7 (X_2 - K_5)_+^1 + \hat{\beta}_8 (X_2 - K_6)_+^1 + \hat{\beta}_9 (X_2 - K_7)_+^1 + \hat{\beta}_{10} (X_2 - K_8)_+^1 + \hat{\beta}_{11} X_3 + \hat{\beta}_{12} (X_3 - K_9)_+^1 + \hat{\beta}_{13} (X_3 - K_{10})_+^1 + \hat{\beta}_{14} (X_3 - K_{11})_+^1 + \hat{\beta}_{15} (X_3 - K_{12})_+^1 + \hat{\beta}_{16} X_4 + \hat{\beta}_{17} (X_4 - K_{13})_+^1 + \hat{\beta}_{18} (X_4 - K_{14})_+^1 + \hat{\beta}_{19} (X_4 - K_{15})_+^1 + \hat{\beta}_{20} (X_4 - K_{16})_+^1 + \hat{\beta}_{21} X_5 + \hat{\beta}_{22} (X_5 - K_{17})_+^1 + \hat{\beta}_{23} (X_5 - K_{18})_+^1 + \hat{\beta}_{24} (X_5 - K_{19})_+^1 + \hat{\beta}_{25} (X_5 - K_{20})_+^1 + \hat{\beta}_{26} X_6 + \hat{\beta}_{27} (X_6 - K_{21})_+^1 + \hat{\beta}_{28} (X_6 - K_{22})_+^1 + \hat{\beta}_{29} (X_6 - K_{23})_+^1 + \hat{\beta}_{30} (X_6 - K_{24})_+^1 \quad (8)$$

dengan \hat{Y} adalah rata-rata lama sekolah masing-masing provinsi di Indonesia tahun 2019, X_1 adalah persentase penduduk miskin, X_2 adalah proporsi disabilitas anak, X_3 adalah angka melek huruf, X_4 adalah angka putus sekolah jenjang SD, X_5 adalah angka putus sekolah jenjang SMP dan X_6 adalah angka putus sekolah jenjang SMA.

Estimasi parameter *spline* diperoleh menggunakan *software* R dan disajikan dalam Tabel 3. Parameter yang digunakan

dilambangkan dengan beta (β) dan diperoleh sebanyak 31.

Estimasi parameter beserta titik *knot* optimal yang telah diperoleh kemudian disubstitusikan dalam persamaan 8. Adapun estimasi model *spline* yang diperoleh dapat dilihat pada persamaan (9).

Tabel 3. Estimasi Parameter Model

Parameter	Nilai Estimasi	Parameter	Nilai Estimasi
$\hat{\beta}_0$	13,36	$\hat{\beta}_{16}$	-1,89
$\hat{\beta}_1$	-0,20	$\hat{\beta}_{17}$	-1,80
$\hat{\beta}_2$	0,19	$\hat{\beta}_{18}$	4,72
$\hat{\beta}_3$	-0,02	$\hat{\beta}_{19}$	6,19
$\hat{\beta}_4$	-0,10	$\hat{\beta}_{20}$	-9,34
$\hat{\beta}_5$	0,18	$\hat{\beta}_{21}$	0,93
$\hat{\beta}_6$	-0,32	$\hat{\beta}_{22}$	-3,89
$\hat{\beta}_7$	2,81	$\hat{\beta}_{23}$	4,08
$\hat{\beta}_8$	-5,67	$\hat{\beta}_{24}$	-0,66
$\hat{\beta}_9$	5,1	$\hat{\beta}_{25}$	-0,75
$\hat{\beta}_{10}$	-2,23	$\hat{\beta}_{26}$	2,24
$\hat{\beta}_{11}$	-0,04	$\hat{\beta}_{27}$	-2,66
$\hat{\beta}_{12}$	0,23	$\hat{\beta}_{28}$	-1,71
$\hat{\beta}_{13}$	-0,14	$\hat{\beta}_{29}$	-1,34
$\hat{\beta}_{14}$	0,72	$\hat{\beta}_{30}$	-0,22
$\hat{\beta}_{15}$	-1,27		

$$\hat{Y} = 13,36 - 0,20x_1 + 0,19(x_1 - 6,1)_+^1 - 0,02(x_1 - 7,88)_+^1 - 0,10(x_1 - 12,34)_+^1 + 0,18(x_1 - 14,99)_+^1 - 0,32x_2 + 2,81(x_2 - 2,5)_+^1 - 5,67(x_2 - 3,1)_+^1 + 5,1(x_2 - 3,8)_+^1 - 2,23(x_2 - 4,8)_+^1 - 0,04x_3 + 0,23(x_3 - 93,54)_+^1 - 0,14(x_3 - 97,72)_+^1 + 0,72(x_3 - 98,75)_+^1 - 1,27(x_3 - 99,03)_+^1 - 1,89x_4 - 1,80(x_4 - 0,18)_+^1 + 4,72(x_4 - 0,32)_+^1 + 6,19(x_4 - 0,6)_+^1 - 9,34(x_4 - 0,65)_+^1 + 0,93x_5 - 3,89(x_5 - 0,32)_+^1 + 4,08(x_5 - 0,84)_+^1 - 0,66(x_5 - 1,17)_+^1 - 0,75(x_5 - 1,71)_+^1 + 2,24x_6 - 2,66(x_6 - 0,56)_+^1 + 1,71(x_6 - 1,43)_+^1 - 1,34(x_6 - 2,06)_+^1 - 0,22(x_6 - 3,84)_+^1 \quad (9)$$

Berdasarkan model tersebut diperoleh nilai koefisien determinasi sebesar 0,9991016. Hal tersebut menunjukkan bahwa model yang diperoleh mampu menjelaskan keragaman nilai RLS masing-masing provinsi di Indonesia menggunakan regresi nonparametrik *spline* dengan variabel persentase penduduk miskin, proporsi disabilitas anak, angka melek huruf, angka putus sekolah jenjang SD, angka putus

sekolah jenjang SMP dan angka putus sekolah jenjang SMA sebesar 99,91016% dan sisanya sebesar 0,0008984% dijelaskan oleh variabel lain diluar model ataupun galat.

3.4 Interpretasi Model *Spline*

Interpretasi dari model regresi nonparametrik *spline* linear dengan empat titik *knot* dapat dijabarkan sebagai berikut :

1. Apabila X_2, X_3, X_4, X_5 dan X_6 konstan maka pengaruh persentase kemiskinan (X_1) terhadap RLS (Y) yaitu

$$\hat{Y} = 13,36 - 0,20X_1 + 0,19(X_1 - 6,1)_+^1 - 0,02(X_1 - 7,88)_+^1 - 0,10(X_1 - 12,34)_+^1 + 0,18(X_1 - 14,99)_+^1 \quad (10)$$

$$\hat{Y} = \begin{cases} 13,36 - 0,20X_1 & ; X_1 < 6,1 \\ 12,201 - 0,01X_1 & ; 6,1 \leq X_1 < 7,88 \\ 12,3586 - 0,03X_1 & ; 7,88 \leq X_1 < 12,34 \\ 13,5926 - 0,13X_1 & ; 12,34 \leq X_1 < 14,99 \\ 10,8944 + 0,05X_1 & ; X_1 \geq 14,99 \end{cases} \quad (11)$$

Apabila persentase kemiskinan (X_1) suatu provinsi berada di bawah angka 6,1% maka peningkatan persentase kemiskinan (X_1) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,20 tahun. Apabila persentase kemiskinan (X_1) suatu provinsi berada pada interval 6,1% sampai 7,88% maka peningkatan persentase kemiskinan (X_1) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,01 tahun. Apabila persentase kemiskinan (X_1) suatu provinsi berada pada interval 7,88% sampai 12,34% maka peningkatan persentase kemiskinan (X_1) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,03 tahun. Apabila persentase kemiskinan (X_1) suatu provinsi berada pada interval 12,34% sampai 14,99% maka peningkatan persentase kemiskinan (X_1) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,13 tahun. Apabila persentase kemiskinan (X_1) suatu provinsi berada di atas angka 14,99% maka peningkatan persentase kemiskinan (X_1) sebesar satu persen akan mengakibatkan peningkatan RLS (Y) sebesar 0,05 tahun.

2. Apabila X_1, X_3, X_4, X_5 dan X_6 konstan maka pengaruh proporsi disabilitas anak (X_2) terhadap RLS (Y) yaitu

$$\hat{Y} = 13,36 - 0,32X_2 + 2,81(X_2 - 2,5)_+^1 \quad (12)$$

$$\begin{aligned} & -5,67(X_2 - 3,1)_+^1 \\ & + 5,1(X_2 - 3,8)_+^1 \\ & - 2,23(X_2 - 4,8)_+^1 \end{aligned}$$

$$\hat{Y} = \begin{cases} 13,36 - 0,32X_2 & ; X_2 < 2,5 \\ 6,335 + 2,49X_2 & ; 2,5 \leq X_2 < 3,1 \\ 23,912 - 3,18X_2 & ; 3,1 \leq X_2 < 3,8 \\ 4,532 + 1,92X_2 & ; 3,8 \leq X_2 < 4,8 \\ 15,236 - 0,31X_2 & ; X_2 \geq 4,8 \end{cases} \quad (13)$$

Apabila proporsi disabilitas anak (X_2) suatu provinsi berada di bawah angka 2,5% maka peningkatan proporsi disabilitas anak (X_2) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,32 tahun. Apabila proporsi disabilitas anak (X_2) suatu provinsi berada pada interval 2,5% sampai 3,1% maka peningkatan proporsi disabilitas anak (X_2) sebesar satu persen akan mengakibatkan peningkatan RLS (Y) sebesar 2,49 tahun. Apabila proporsi disabilitas anak (X_2) suatu provinsi berada pada interval 3,1% sampai 3,8% maka peningkatan proporsi disabilitas anak (X_2) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 3,18 tahun. Apabila proporsi disabilitas anak (X_2) suatu provinsi berada pada interval 3,8% sampai 4,8% maka peningkatan proporsi disabilitas anak (X_2) sebesar satu persen akan mengakibatkan peningkatan RLS (Y) sebesar 1,92 tahun. Apabila proporsi disabilitas anak (X_2) suatu provinsi berada di atas angka 4,8% maka peningkatan proporsi disabilitas anak (X_2) sebesar satu persen akan mengakibatkan penurunan RLS (Y) sebesar 0,31 tahun.

Begitu pula untuk hubungan X_3, X_4, X_5, X_6 terhadap rata-rata lama sekolah (Y).

4. SIMPULAN DAN SARAN

4.1 Simpulan

Model regresi terbaik yang diperoleh dalam pemodelan RLS di Indonesia dengan variabel persentase penduduk miskin, proporsi disabilitas anak, angka melek huruf, angka putus sekolah jenjang SD, angka putus sekolah jenjang SMP dan angka putus sekolah jenjang SMA yaitu menggunakan orde linear (orde 2) dengan empat titik *knot*. Nilai koefisien determinasi yang diperoleh sebesar 0,999 dengan nilai GCV yaitu 0,079.

4.2 Saran

Saran yang dapat penulis berikan untuk penelitian selanjutnya yaitu sebaiknya dapat menyederhanakan model yang diperoleh agar sesuai dengan *principle of parsimony model* atau prinsip kesederhanaan model tanpa mengurangi nilai koefisien determinasi yang diperoleh.

DAFTAR PUSTAKA

- Astiti, D. A. W., Sumarjaya, I., & Susilawati, M. (2016). Analisis Regresi Nonparametrik Spline Multivariat untuk Pemodelan Indikator Kemiskinan di Indonesia. *E-Jurnal Matematika*, 5(3), 111-116.
- BPS. (2020). Potret Pendidikan Indonesia 2019. *Statistik Pendidikan*.
- Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing Second edition*. New York: Marcel Dekker.
- Griggs, W. (2013). *Penalized Spline Regression and Its Applications*. Washington D.C.: Whitman College.
- Kurtner, M., & Nachtsheim, C. (2004). *Applied Linear Statistical Models*. New York: The McGraw-Hill Companies.
- Prahtama, A., Utama, T., Caraka, R., & Zumrohtuliyosi, D. (2014). Pemodelan Inflasi Berdasarkan Harga-Harga Pangan Menggunakan Spline Multivariabel. *Media Statistika*. 7(2), 89-94.