

# Pemodelan APM Usia SMA dan Sederajat di Provinsi Papua Menggunakan Metode Regresi Nonparametrik *Spline* Tahun 2019

Kornelia Zenitha Vyamili dan Sri Pingit Wulandari  
Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember (ITS)  
*e-mail*: sripingitwulandari@gmail.com

**Abstrak**—Pendidikan merupakan suatu hal yang sangat penting bagi setiap manusia agar menjadi sumber daya manusia yang berkualitas. Angka Partisipasi Murni (APM) adalah salah satu indikator yang dapat menunjukkan tingkat pendidikan karena APM dapat mengukur daya serap sistem pendidikan terhadap penduduk usia sekolah. APM usia SMA dan sederajat di Provinsi Papua sebesar 44,32% adalah yang paling rendah di antara 34 provinsi di Indonesia. Sebagai upaya untuk meningkatkan APM di Provinsi Papua, perlu dilakukan pemodelan APM usia SMA dan sederajat sehingga dapat diketahui faktor-faktor apa saja yang mempengaruhinya. Metode regresi nonparametrik *spline* sangat fleksibel dalam pemodelan dan sesuai digunakan pada penelitian di bidang sosial dan kependudukan, sehingga dapat digunakan untuk memodelkan APM usia SMA dan sederajat di Provinsi Papua. Hasil penelitian menunjukkan bahwa persentase penduduk miskin, rasio murid-guru, rasio murid-sekolah, angka melek huruf, dan harapan lama sekolah berpengaruh signifikan terhadap APM usia SMA dan sederajat. Model APM usia SMA dan sederajat di Provinsi Papua memiliki koefisien determinasi ( $R^2$ ) sebesar 98,84% dengan MSE sebesar 16,786.

**Kata Kunci**—Angka Partisipasi Murni, Provinsi Papua, Regresi Nonparametrik, *Spline*.

## I. PENDAHULUAN

PAPUA merupakan provinsi terbesar di Indonesia dengan wilayah seluas 316.553,07 km<sup>2</sup>. Potensi sumber daya alam di Provinsi Papua bernilai ekonomis dan strategis yang dapat dimanfaatkan secara optimal jika didukung oleh generasi muda yang berkualitas. Kualitas sumber daya manusia dapat dilihat melalui tingkat pendidikan. Provinsi Papua memiliki tingkat pendidikan yang rendah jika diukur berdasarkan Angka Partisipasi Murni (APM) usia SMA dan sederajat tahun 2019 yaitu sebesar 44,32%, dimana angka tersebut merupakan nilai paling rendah di antara 33 provinsi lainnya di Indonesia. Adapun target nasional APM usia SMA dan sederajat tahun 2019 menurut Rancangan Teknokratik RPJMN 2015 – 2019 adalah sebesar 67,5%.

Untuk mengejar ketertinggalan tingkat pendidikan Provinsi Papua yang dapat disebabkan oleh beberapa faktor, diperlukan penelitian mengenai faktor-faktor yang memengaruhi tingkat pendidikan dengan melakukan pemodelan APM usia SMA dan sederajat di Provinsi Papua. Persentase penduduk miskin, rasio murid-guru, rasio murid-sekolah, angka melek huruf, dan harapan lama sekolah digunakan sebagai variabel prediktor dalam pemodelan APM usia SMA dan sederajat di Provinsi Papua.

Regresi nonparametrik digunakan untuk mengetahui pola hubungan antara variabel respon dengan variabel prediktor tanpa informasi sebelumnya mengenai bentuk fungsi regresinya. Regresi nonparametrik juga dapat digunakan dengan

mengabaikan asumsi-asumsi yang melandasi metode regresi parametrik sehingga sangat fleksibel dalam memodelkan pola data. Fungsi *spline* yang mulus dan fleksibel memungkinkan untuk menyesuaikan diri secara efektif terhadap karakteristik dari fungsi suatu data [1]. APM usia SMA dan sederajat di Provinsi Papua dan faktor-faktor yang diduga memengaruhinya merupakan persoalan sosial yang tidak memiliki informasi mengenai bentuk fungsi regresinya. Sehingga, metode regresi nonparametrik *spline* dapat digunakan untuk melakukan pemodelan APM usia SMA dan sederajat di Provinsi Papua.

## II. STUDI LITERATUR

### A. Regresi Nonparametrik *Spline*

Regresi nonparametrik digunakan ketika informasi sebelumnya mengenai bentuk fungsi atau kurva regresi suatu data tidak diketahui. Kelebihan dari regresi ini adalah data dibiarkan mencari bentuk estimasi dari kurva regresinya tanpa dipengaruhi oleh subyektivitas peneliti. Model umum dari regresi nonparametrik untuk variabel prediktor lebih dari satu dengan  $n$  pengamatan ditunjukkan pada persamaan sebagai berikut:

$$y_i = \sum_{j=1}^p f(x_{ji}) + \varepsilon_i; i = 1, 2, 3, \dots, n \quad (1)$$

dengan  $y_i$  adalah variabel respon,  $x_i$  adalah variabel prediktor,  $f(x_{ji})$  adalah fungsi regresi yang tidak diketahui bentuknya, dan  $\varepsilon_i$  adalah galat acak.

Salah satu pendekatan untuk mengestimasi fungsi regresi nonparametrik yang tidak diketahui bentuknya adalah *spline*. *Spline* merupakan model polinom tersegmen yang mulus dan lebih fleksibel dari polinom biasa, sehingga memungkinkan untuk menyesuaikan diri secara efektif terhadap karakteristik dari fungsi suatu data [1]. *Spline* sangat bergantung pada titik knot yang merupakan titik perpaduan bersama dimana terjadi pola perubahan perilaku dari suatu fungsi pada interval yang berbeda [2]. Fungsi *spline* variabel prediktor ke- $j$  dengan orde  $m$  dan titik knot sebanyak  $r$  untuk  $f(x_{ji})$  pada persamaan (1) dapat dinyatakan sebagai berikut:

$$f(x_{ji}) = \sum_{m=0}^q \beta_{mj} x_{ji}^m + \sum_{k=1}^r \beta_{(q+k)j} (x_{ji} - K_{kj})_+^q \quad (2)$$

dimana  $k$  merupakan banyaknya titik knot.  $(x_{ji} - K_{kj})_+^q$  menyatakan fungsi potongan (*truncated*) yang dijabarkan sebagai berikut:

$$(x_{ji} - K_{kj})_+^q = \begin{cases} (x_{ji} - K_{kj})^q, & x_{ji} - K_{kj} \geq 0 \\ 0, & x_{ji} - K_{kj} < 0 \end{cases} \quad (3)$$

Tabel 1.  
Daerah pengujian durbin watson

Kesimpulan	Daerah Pengujian
Terdapat autokorelasi positif	$d < d_L$
Tidak dapat disimpulkan	$d_L < d < d_U$ atau $4-d_U < d < 4-d_L$
Tidak terdapat autokorelasi	$d_U < d < 4-d_U$
Terdapat autokorelasi negatif	$4-d_L < d$

Tabel 2.  
Variabel penelitian

Variabel	Keterangan	Skala
Y	APM Usia SMA dan Sederajat	Rasio
X <sub>1</sub>	Persentase Penduduk Miskin	Rasio
X <sub>2</sub>	Rasio Murid-Guru	Rasio
X <sub>3</sub>	Rasio Murid-Sekolah	Rasio
X <sub>4</sub>	Angka Melek Huruf	Rasio
X <sub>5</sub>	Harapan Lama Sekolah	Rasio

Tabel 3.  
Struktur data

Kabupaten/Kota	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
1	Y <sub>1</sub>	X <sub>11</sub>	X <sub>21</sub>	X <sub>31</sub>	X <sub>41</sub>	X <sub>51</sub>
2	Y <sub>2</sub>	X <sub>12</sub>	X <sub>22</sub>	X <sub>32</sub>	X <sub>42</sub>	X <sub>52</sub>
3	Y <sub>3</sub>	X <sub>13</sub>	X <sub>23</sub>	X <sub>33</sub>	X <sub>43</sub>	X <sub>53</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	Y <sub>29</sub>	X <sub>129</sub>	X <sub>229</sub>	X <sub>329</sub>	X <sub>429</sub>	X <sub>529</sub>

Pendugaan terhadap vektor parameter  $\beta$  dilakukan dengan menggunakan metode *least square* sebagai berikut:

$$\hat{\beta} = (X^T T)^{-1} X^T Y \tag{4}$$

Estimator *spline* yang baik diperoleh menggunakan titik knot ( $K$ ) yang optimal. Pemilihan titik knot yang optimal sangatlah penting dalam menentukan model terbaik dalam regresi *spline*. Pemilihan estimator regresi *spline* terbaik di antara model-model yang didapatkan dilihat berdasarkan kriteria *Generalized Cross Validation* (GCV) yang minimum.

GCV merupakan suatu metode yang populer digunakan untuk memilih parameter *smoothing* dalam *spline*. Kelebihan yang dimiliki oleh GCV adalah optimal secara asimtotik, invarian terhadap transformasi dan tidak memerlukan informasi terhadap varian dalam menghitung nilai yang bersifat sederhana [3]. Nilai GCV didapatkan sebagaimana pada persamaan berikut:

$$GCV = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n^{-1} tr [I - H])^2} \tag{5}$$

dengan  $H = X(X^T X)^{-1} X^T$ ,  $n$  adalah banyaknya pengamatan, dan  $I$  adalah matriks identitas.

**B. Pengujian Parameter Model Regresi**

Pengujian parameter dilakukan untuk mengetahui apakah suatu variabel memberikan pengaruh yang signifikan dalam model. Pengujian parameter dapat dilakukan secara serentak dan individu.

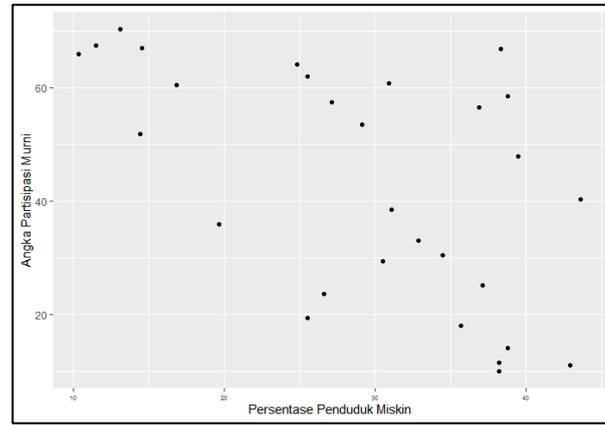
**1) Pengujian Serentak**

Pengujian parameter secara serentak atau uji  $F$  merupakan uji signifikansi model secara keseluruhan untuk mengetahui apakah variabel prediktor yang dimasukkan ke dalam model memberikan pengaruh signifikan terhadap variabel respon secara bersama-sama. Hipotesis yang digunakan dalam pengujian parameter secara serentak adalah sebagai berikut.

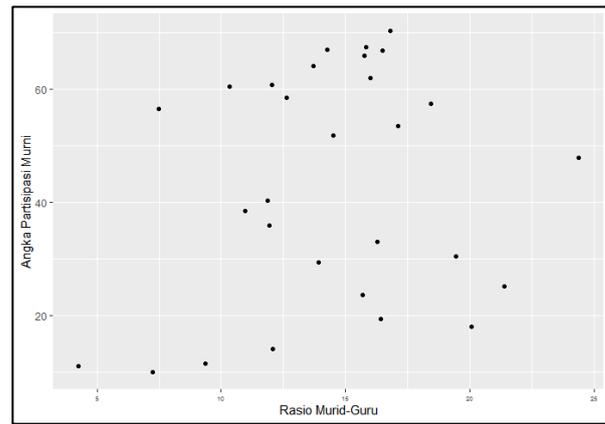
$$H_0: \beta_{11} = \beta_{21} = \dots = \beta_{(q+r)p} = 0$$

$$H_1: \text{minimal ada satu } \beta_{lj} \neq 0; \text{ dimana } l = 1, 2, 3, \dots, q + r \text{ dan } j = 1, 2, 3, \dots, p.$$

Statistik uji yang digunakan adalah  $F_{hit}$  sebagaimana yang



Gambar 1. Scatter plot persentase penduduk miskin vs APM.



Gambar 2. Scatter plot rasio murid-guru vs APM.

ditunjukkan pada persamaan berikut:

$$F_{hit} = \frac{MS_{regression}}{MS_{error}} \tag{6}$$

Daerah penolakan yang digunakan adalah dengan membandingkan nilai  $F_{hit}$  dengan nilai  $F_{\alpha(p(q+r), n-p(q+r)-1)}$ . Dengan menggunakan taraf signifikan sebesar  $\alpha$ , maka keputusan bernilai tolak  $H_0$  jika  $F_{hit} > F_{\alpha(p(q+r), n-p(q+r)-1)}$  yang berarti bahwa model signifikan [4].

**2) Pengujian Individu**

Pengujian parameter secara individu atau uji  $t$  merupakan uji signifikansi masing-masing variabel prediktor untuk mengetahui apakah variabel tersebut memberikan pengaruh signifikan terhadap variabel respon secara individu. Berikut adalah hipotesis yang digunakan dalam pengujian.

$$H_0: \beta_{lj} = 0$$

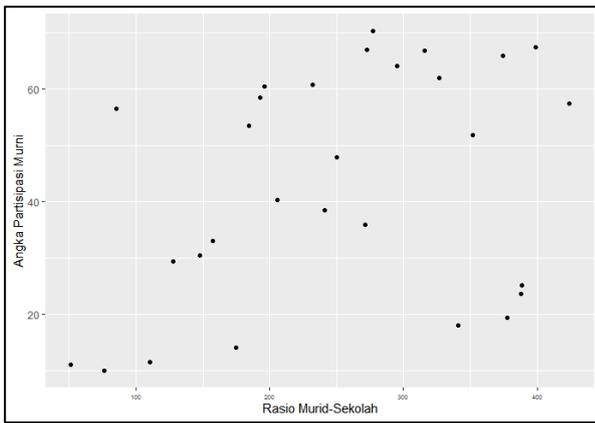
$$H_1: \beta_{lj} \neq 0; \text{ dimana } l = 1, 2, 3, \dots, q + r \text{ dan } j = 1, 2, 3, \dots, p.$$

Statistik uji yang digunakan ditunjukkan pada persamaan sebagai berikut:

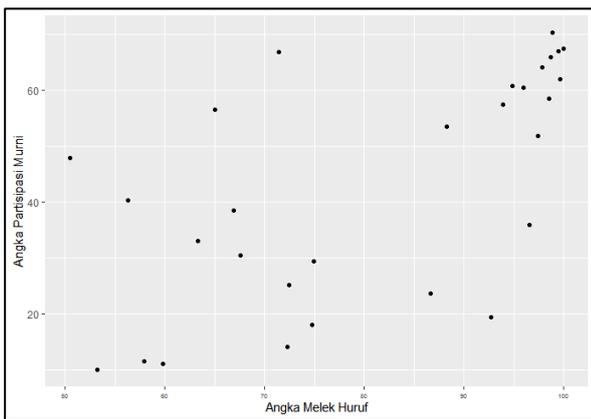
$$t_{hit} = \frac{\hat{\beta}_{lj}}{SE(\hat{\beta}_{lj})} \tag{7}$$

$$\text{dengan } SE(\hat{\beta}_{lj}) = \sqrt{MSE(X^T X)^{-1}}$$

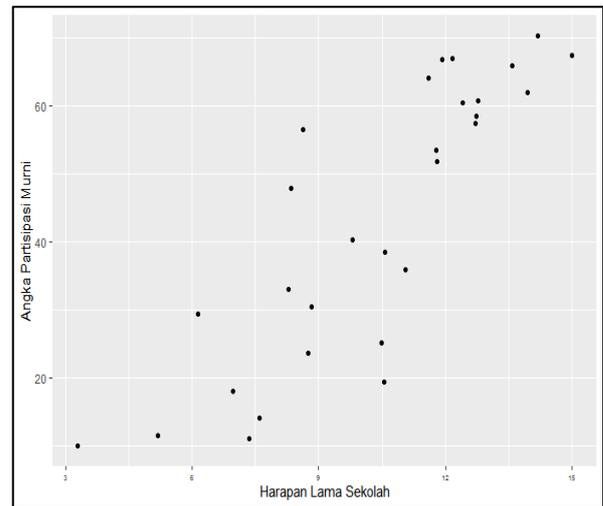
Daerah penolakan yang digunakan adalah dengan membandingkan nilai  $t_{hit}$  dengan  $t_{\frac{\alpha}{2}(n-p(q+r)-1)}$ . Dengan menggunakan taraf signifikan sebesar  $\alpha$ , maka keputusan bernilai tolak  $H_0$  jika  $|t_{hit}| > t_{\frac{\alpha}{2}(n-p(q+r)-1)}$  yang berarti bahwa parameter berpengaruh secara signifikan terhadap



Gambar 3. Scatter plot rasio murid-sekolah vs APM.



Gambar 4. Scatter plot angka melek huruf vs APM.



Gambar 5. Scatter plot harapan lama sekolah vs APM.

model [4].

C. Pengujian Asumsi Residual

Pengujian asumsi residual dilakukan untuk mengetahui apakah residual telah memenuhi asumsi regresi. Asumsi regresi meliputi asumsi identik, independen, dan berdistribusi normal.

1) Asumsi Residual Identik

Pengujian yang dapat digunakan untuk mengetahui apakah residual telah memenuhi asumsi identik adalah Uji Park, Uji Glejser, dan Uji Spearman's Correlation. Uji Glejser dilakukan dengan meregresikan variabel-variabel bebas terhadap nilai absolut residualnya [5]. Hipotesis yang digunakan dalam Uji Glejser adalah sebagai berikut:

- $H_0$  : residual identik (tidak terdapat heteroskedastisitas)
- $H_1$  : residual tidak identik (terdapat heteroskedastisitas)

Statistik uji yang digunakan ditunjukkan pada persamaan sebagai berikut:

$$F_{hit} = \frac{MS_{regression}}{MS_{error}} \tag{8}$$

Dengan menggunakan taraf signifikan sebesar  $\alpha$ , maka keputusan bernilai tolak  $H_0$  jika nilai  $F_{hit}$  lebih dari nilai  $F_{\alpha(p(q+r), n-p(q+r)-1)}$  atau apabila  $p$ -value kurang dari taraf signifikan.

2) Asumsi Residual Independen

Durbin Watson, Lagrange Multiplier, dan Run Test merupakan uji-uji yang dapat digunakan untuk mendeteksi adanya autokorelasi (korelasi antar residual) atau untuk mengetahui apakah residual telah memenuhi asumsi independen [5]. Berikut adalah hipotesis yang digunakan

dalam Uji Durbin Watson:

- $H_0$  : residual independen (tidak terdapat autokorelasi)
- $H_1$  : residual tidak independen (terdapat autokorelasi)

Statistik uji yang digunakan adalah  $d$  sebagaimana ditunjukkan pada persamaan berikut:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \tag{9}$$

Dengan menggunakan taraf signifikan sebesar  $\alpha$ , maka keputusan bernilai tolak  $H_0$  jika nilai  $d$  lebih dari nilai  $d_L$  atau  $p$ -value kurang dari taraf signifikan. Adapun kesimpulan dari masing-masing daerah pengujian ditunjukkan pada Tabel 1.

3) Asumsi Residual Berdistribusi Normal

Suatu residual dapat dikatakan berdistribusi normal apabila memiliki  $(\sigma^2, \mu)$  dimana varian bersifat konstan dan mean bernilai nol. Pengujian yang dapat digunakan untuk mengetahui apakah residual telah memenuhi asumsi distribusi normal adalah Uji Kolmogorov-Smirnov, Uji Chi-Square, dan Uji Liliefors [6]. Berikut adalah hipotesis yang digunakan dalam Uji Kolmogorov-Smirnov:

- $H_0$  : residual berdistribusi normal
- $H_1$  : residual tidak berdistribusi normal

Statistik uji yang digunakan dalam Uji Kolmogorov-Smirnov ditunjukkan pada persamaan sebagai berikut:

$$D = \text{Sup} |F_n(x) - F_0(x)| \tag{10}$$

dimana  $F_n(x)$  merupakan nilai distribusi kumulatif dan  $F_0(x)$  menunjukkan nilai distribusi kumulatif yang diterapkan.

Dengan menggunakan taraf signifikan sebesar  $\alpha$ , maka keputusan bernilai tolak  $H_0$  jika nilai  $D$  lebih dari nilai  $D_{\alpha, n}$  atau  $p$ -value kurang dari taraf signifikan.

Tabel 4.

Nilai GCV minimum satu, dua, tiga, dan kombinasi titik knot

GCV minimum	Titik Knot Optimum				
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
146,579	15,107	7,125	104,318	57,580	4,963
161,874	11,539	4,969	64,329	52,278	3,708
	13,918	6,407	90,988	55,813	4,545
60,850	32,946	17,908	304,258	84,093	11,236
	34,136	18,627	317,587	85,860	11,654
	40,082	22,222	384,234	94,698	13,745
60,850	32,946	17,908	304,258	84,093	11,236
	34,136	18,627	317,587	85,860	11,654
	40,082	22,222	384,234	94,698	13,745

Tabel 5.

Estimasi parameter model regresi nonparametrik <i>Spline</i>					
Variabel	Parameter	Estimasi	Variabel	Parameter	Estimasi
Const.	$\beta_0$	-37,516	X <sub>3</sub>	$\beta_{33}$	0,990
	$\beta_{11}$	0,087		$\beta_{43}$	2,288
	$\beta_{12}$	89,205		$\beta_{14}$	0,480
X <sub>1</sub>	$\beta_{13}$	-114,839	X <sub>4</sub>	$\beta_{24}$	-12,343
	$\beta_{14}$	41,335		$\beta_{34}$	12,280
	$\beta_{12}$	0,0617		$\beta_{44}$	1,223
X <sub>2</sub>	$\beta_{22}$	-165,631	X <sub>5</sub>	$\beta_{15}$	1,551
	$\beta_{32}$	190,940		$\beta_{25}$	62,457
	$\beta_{42}$	7,866		$\beta_{35}$	-58,184
X <sub>3</sub>	$\beta_{13}$	0,073		$\beta_{45}$	-32,755
	$\beta_{23}$	-1,093			

Tabel 6.

Penguji parameter secara serentak			
F	F <sub>0,05(20;8)</sub>	p-value	Keputusan
34,121	3,150	0,000	Tolak H <sub>0</sub>

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder tahun 2019 yang diambil dari *website* resmi Badan Pusat Statistik Provinsi Papua yaitu [www.papua.bps.go.id](http://www.papua.bps.go.id).

#### B. Variabel Penelitian

Penelitian ini terdiri dari variabel respon berupa Angka Partisipasi Murni (APM) usia SMA dan sederajat dan variabel prediktor berupa faktor-faktor yang diduga mempengaruhi APM usia SMA dan sederajat. Variabel penelitian dapat dilihat pada Tabel 2. Struktur data dengan unit penelitian 29 kabupaten/kota di Provinsi Papua dapat dilihat pada Tabel 3.

#### C. Langkah Analisis

Langkah analisis yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan karakteristik Provinsi Papua berdasarkan Angka Partisipasi Murni (APM) usia SMA dan sederajat beserta faktor-faktor yang diduga memengaruhinya.
2. Mengidentifikasi pola hubungan antara APM usia SMA dan sederajat dengan faktor-faktor yang diduga memengaruhinya menggunakan *scatter plot*.
3. Menentukan titik knot optimum berdasarkan nilai *Generalized Cross Validation* (GCV) minimum.
4. Mengestimasi parameter model regresi nonparametrik *spline* linear dan membentuk model APM usia SMA dan sederajat berdasarkan faktor-faktor yang memengaruhinya.
5. Menguji parameter model regresi nonparametrik *spline* secara serentak dan individu.
6. Menginterpretasi model yang terbentuk.
7. Menguji asumsi residual identik, independen, dan berdistribusi normal.
8. Menarik kesimpulan.

Tabel 7.

Penguji parameter secara individu				
Var.	Parameter	t	p-value	Keputusan
Const.	$\beta_0$	2,161	0,039	Tolak H <sub>0</sub>
	$\beta_{11}$	0,358	0,723	Gagal Tolak H <sub>0</sub>
	$\beta_{12}$	7,514	0,000	Tolak H <sub>0</sub>
X <sub>1</sub>	$\beta_{13}$	7,634	0,000	Tolak H <sub>0</sub>
	$\beta_{14}$	7,194	0,000	Tolak H <sub>0</sub>
	$\beta_{12}$	1,183	0,247	Gagal Tolak H <sub>0</sub>
X <sub>2</sub>	$\beta_{22}$	6,845	0,000	Tolak H <sub>0</sub>
	$\beta_{32}$	6,602	0,000	Tolak H <sub>0</sub>
	$\beta_{42}$	0,938	0,356	Gagal Tolak H <sub>0</sub>
X <sub>3</sub>	$\beta_{13}$	2,574	0,016	Tolak H <sub>0</sub>
	$\beta_{23}$	2,677	0,012	Tolak H <sub>0</sub>
	$\beta_{33}$	2,055	0,049	Tolak H <sub>0</sub>
X <sub>4</sub>	$\beta_{43}$	5,423	0,000	Tolak H <sub>0</sub>
	$\beta_{14}$	1,880	0,071	Gagal Tolak H <sub>0</sub>
	$\beta_{24}$	3,137	0,004	Tolak H <sub>0</sub>
X <sub>5</sub>	$\beta_{34}$	2,905	0,007	Tolak H <sub>0</sub>
	$\beta_{44}$	0,858	0,398	Gagal Tolak H <sub>0</sub>
	$\beta_{15}$	1,715	0,097	Gagal Tolak H <sub>0</sub>
X <sub>5</sub>	$\beta_{25}$	6,039	0,000	Tolak H <sub>0</sub>
	$\beta_{35}$	5,381	0,000	Tolak H <sub>0</sub>
	$\beta_{45}$	4,629	0,000	Tolak H <sub>0</sub>

### IV. ANALISIS DAN PEMBAHASAN

#### A. Pola Hubungan Antara APM Usia SMA dan Sederajat dengan Faktor-Faktor yang Diduga Memengaruhinya

Identifikasi pola hubungan dilakukan untuk melihat apakah metode regresi nonparametrik tepat digunakan dalam pemodelan. Pola hubungan antara APM usia SMA dan sederajat dengan persentase penduduk miskin, rasio murid-guru, rasio murid-sekolah, angka melek huruf, dan harapan lama sekolah yang diidentifikasi menggunakan *scatter plot* ditunjukkan pada Gambar 1 hingga Gambar 5.

Plot data pada Gambar 1 yang menyebar secara acak menunjukkan bahwa hubungan antara APM usia SMA dan sederajat dengan persentase penduduk miskin tidak membentuk suatu pola tertentu, maka estimasi model APM usia SMA dan sederajat dapat dilakukan menggunakan metode regresi nonparametrik.

Hubungan antara APM usia SMA dan sederajat dengan rasio murid-guru tidak membentuk suatu pola tertentu karena plot data pada Gambar 2 menyebar secara acak, maka metode regresi nonparametrik dapat digunakan dalam mengestimasi model APM usia SMA dan sederajat.

Plot data pada Gambar 3 yang menyebar secara acak menunjukkan bahwa hubungan antara APM usia SMA dan sederajat dengan rasio murid-sekolah tidak membentuk suatu pola tertentu, maka estimasi model APM usia SMA dan sederajat dapat dilakukan menggunakan metode regresi nonparametrik.

Hubungan antara APM usia SMA dan sederajat dengan angka melek huruf tidak membentuk suatu pola tertentu karena plot data pada Gambar 4 menyebar secara acak, maka

Tabel 8.  
Uji glejser

F	F <sub>0,05(20;8)</sub>	p-value
1,225	3,150	0,403

Tabel 9.  
Uji durbin watson

d <sub>L</sub>	d	d <sub>U</sub>	p-value
1,0497	2,561	1,8409	0,64

Tabel 10.  
Uji kolmogorov-smirnov

KS	KS <sub>(0,05;29)</sub>	p-value
0,124	0,246	0,722

metode regresi nonparametrik dapat digunakan dalam mengestimasi model APM usia SMA dan sederajat.

Plot data pada Gambar 5 yang menyebar secara acak menunjukkan bahwa hubungan antara APM usia SMA dan sederajat dengan harapan lama sekolah tidak membentuk suatu pola tertentu, maka estimasi model APM usia SMA dan sederajat dapat dilakukan menggunakan metode regresi nonparametrik.

**B. Pemodelan APM Usia SMA dan Sederajat**

Pemodelan Angka Partisipasi Murni (APM) usia SMA dan sederajat menggunakan regresi nonparametrik *spline* dilakukan dengan pemilihan titik knot optimum berdasarkan nilai *Generalized Cross Validation* (GCV) minimum, pengestimasian parameter, pembentukan model, pengujian signifikansi parameter secara serentak dan individu, serta pengujian asumsi residual.

**1) Pemilihan Titik Knot Optimum**

Estimator *spline* yang baik diperoleh dari titik knot optimum yang dipilih berdasarkan nilai GCV minimum. Pemodelan dilakukan dengan menggunakan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot. Nilai GCV minimum ditunjukkan pada Tabel 4.

Tabel 4 menunjukkan bahwa model dengan menggunakan tiga titik knot atau kombinasi knot 3, 3, 3, 3 terpilih menjadi model regresi nonparametrik *spline* yang paling baik karena memiliki nilai GCV paling minimum diantara keempat model lainnya yaitu sebesar 60,85.

**2) Estimasi Parameter Model Regresi Nonparametrik Spline**

Estimasi parameter model regresi nonparametrik *spline* ditunjukkan pada Tabel 5.

**3) Pengujian Signifikansi Parameter Serentak**

Uji signifikansi parameter secara serentak digunakan untuk mengetahui apakah parameter model regresi nonparametrik *spline* berpengaruh signifikan terhadap Angka Partisipasi Murni (APM) usia SMA dan sederajat secara serentak.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_5 = 0$$

$$H_1 : \text{minimal ada satu } \beta_{lj} \neq 0; l = 1, 2, 3, 4; j = 1, 2, 3, 4, 5$$

Hasil pengujian signifikansi parameter model regresi nonparametrik *spline* secara serentak dapat dilihat pada Tabel 6.

Tabel 6 menunjukkan bahwa dengan menggunakan taraf signifikan sebesar 5%, didapatkan keputusan tolak H<sub>0</sub> karena nilai F sebesar 34,121 lebih besar dari F<sub>0,05(20;8)</sub> sebesar 3,150 dan diperkuat oleh p-value sebesar 0,000 kurang dari taraf

signifikan sebesar 0,05. Artinya, minimal terdapat satu parameter yang berpengaruh signifikan terhadap APM usia SMA dan sederajat.

**4) Pengujian Signifikansi Parameter Individu**

Uji signifikansi parameter secara individu digunakan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap Angka Partisipasi Murni (APM) usia SMA dan sederajat.

Hipotesis:

$$H_0 : \beta_{lj} = 0; l = 1, 2, 3, 4; j = 1, 2, 3, 4, 5$$

$$H_1 : \beta_{lj} \neq 0; l = 1, 2, 3, 4; j = 1, 2, 3, 4, 5$$

Hasil pengujian signifikansi parameter model regresi nonparametrik *spline* secara individu dapat dilihat pada Tabel 7.

Hasil uji individu pada Tabel 7 menunjukkan bahwa terdapat 15 parameter yang bernilai tolak H<sub>0</sub> atau berpengaruh signifikan terhadap APM usia SMA dan sederajat karena memiliki p-value kurang dari taraf signifikan sebesar 0,05 dan 6 parameter yang bernilai gagal tolak H<sub>0</sub> atau tidak berpengaruh signifikan terhadap APM usia SMA dan sederajat karena memiliki p-value lebih dari taraf signifikan sebesar 0,05. Meskipun terdapat parameter yang tidak berpengaruh signifikan terhadap model, dapat disimpulkan bahwa semua variabel prediktor berpengaruh signifikan terhadap APM usia SMA dan sederajat.

**5) Interpretasi Model APM Usia SMA dan Sederajat**

Model regresi nonparametrik *spline* linear Angka Partisipasi Murni (APM) usia SMA dan sederajat di Provinsi Papua tahun 2019 adalah sebagai berikut:

$$\hat{y} = -37,516 + 0,087x_1 + 89,205 (x_1 - 32,946)_{\frac{1}{4}} - 114,839 (x_1 - 34,136) + 41,335(x_1 - 40,082)_{\frac{1}{4}} + 0,617 x_2 - 165,631(x_2 - 17,908)_{\frac{1}{4}} + 190,940 (x_2 - 18,627)_{\frac{1}{4}} + 7,866(x_2 - 22,222)_{\frac{1}{4}} + 0,073 x_3 - 1,093 (x_3 - 304,258)_{\frac{1}{4}} + 0,990 (x_3 - 317,587)_{\frac{1}{4}} + 2,288(x_3 - 384,234)_{\frac{1}{4}} + 0,480 x_4 - 12,343 (x_4 - 84,093)_{\frac{1}{4}} + 12,280 (x_4 - 85,860)_{\frac{1}{4}} + 1,222 (x_4 - 94,698)_{\frac{1}{4}} + 1,551 x_5 + 62,457 (x_5 - 11,236)_{\frac{1}{4}} - 58,184 (x_5 - 11,654)_{\frac{1}{4}} - 32,755 (x_5 - 13,745)_{\frac{1}{4}}$$

Model regresi nonparametrik *spline* linear yang terbentuk memiliki nilai R<sup>2</sup> sebesar 98,84%, yang berarti bahwa persentase penduduk miskin, rasio murid-guru, rasio murid-sekolah, angka melek huruf, dan harapan lama sekolah dapat menjelaskan APM usia SMA dan sederajat sebesar 98,84% dan 1,16% sisanya dijelaskan oleh variabel lain di luar model. Interpretasi dari model yang terbentuk adalah sebagai berikut:

Apabila variabel X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, dan X<sub>5</sub> diasumsikan konstan, maka pengaruh variabel X<sub>1</sub> atau persentase penduduk miskin terhadap APM usia SMA dan sederajat adalah sebagai berikut:

$$\hat{y} = 0,087x_1 + 89,205 (x_1 - 32,946)_{\frac{1}{4}} - 114,839 (x_1 - 34,136) + 41,335(x_1 - 40,082)_{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} 0,087x_1 & x_1 < 32,946 \\ 89,292x_1 - 2938,948 & 32,946 \leq x_1 < 34,136 \\ -25,547x_1 + 981,196 & 34,136 \leq x_1 < 40,082 \\ 15,788x_1 - 675,593 & x_1 \geq 40,082 \end{cases}$$

Jika terjadi kenaikan persentase penduduk miskin sebesar 1% pada kondisi persentase penduduk miskin kurang dari 32,95%, maka APM akan meningkat sebesar 0,09%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Mimika, Boven Digoel, Mappi, Asmat, Pegunungan Bintang, Tolikara, Sarmi, Keerom, Waropen, Mamberamo Raya, Dogiyai, dan Kota Jayapura.

Jika terjadi kenaikan persentase penduduk miskin sebesar 1% pada kondisi dimana persentase penduduk miskin berada dalam interval 32,95% hingga 34,14%, maka APM akan meningkat sebesar 82,92%.

Jika terjadi kenaikan persentase penduduk miskin sebesar 1% pada kondisi dimana persentase penduduk miskin berada dalam interval 34,14% hingga 40,08%, maka APM akan menurun sebesar 25,55%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayawijaya, Painai, Puncak Jaya, Yahukimo, Supiori, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, dan Puncak.

Jika terjadi kenaikan persentase penduduk miskin sebesar 1% pada kondisi dimana persentase penduduk miskin lebih dari sama dengan 40,08%, maka APM akan meningkat sebesar 15,79%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Intan Jaya dan Deiyai.

Apabila variabel  $X_1, X_3, X_4,$  dan  $X_5$  diasumsikan konstan, maka pengaruh variabel  $X_2$  atau rasio murid-guru terhadap APM usia SMA dan sederajat adalah sebagai berikut:

$$\hat{y} = 0,617x_2 - 165,631(x_2 - 17,908)_+^{\frac{1}{4}} + 190,940(x_2 - 18,627)_+^{\frac{1}{4}} + 7,866(x_2 - 22,222)_+^{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} 0,617x_2 & x_2 < 17,908 \\ -165,014x_2 + 2966,120 & 17,908 \leq x_2 < 18,627 \\ 25,926x_2 - 590,519 & 18,627 \leq x_2 < 22,222 \\ 33,792x_2 - 765,318 & x_2 \geq 22,222 \end{cases}$$

Jika terjadi kenaikan rasio murid-guru sebesar satu satuan pada kondisi dimana rasio murid-guru kurang dari 17,91, maka APM akan meningkat sebesar 0,62%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Jayawijaya, Jayapura, Nabire, Biak Numfor, Mimika, Boven Digoel, Mappi, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Nduga, Mamberamo Tengah, Puncak, Dogiyai, Intan Jaya, Deiyai, dan Kota Jayapura.

Jika terjadi kenaikan rasio murid-guru sebesar satu satuan pada kondisi dimana rasio murid-guru berada dalam interval 17,91 hingga 18,63, maka APM akan menurun sebesar 165,02%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Kepulauan Yapen.

Jika terjadi kenaikan rasio murid-guru sebesar satu satuan pada kondisi dimana rasio murid-guru berada dalam interval 18,63 hingga 22,22, maka APM akan meningkat sebesar 25,93%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Painai, Puncak Jaya, dan Yalimo.

Jika terjadi kenaikan rasio murid-guru sebesar satu satuan pada kondisi dimana rasio murid-guru lebih dari sama dengan 22,22, maka APM akan meningkat sebesar 33,79%. Wilayah

yang termasuk dalam interval ini adalah Kabupaten Lanny Jaya.

Apabila variabel  $X_1, X_2, X_4,$  dan  $X_5$  diasumsikan konstan, maka pengaruh variabel  $X_3$  atau rasio murid-sekolah terhadap APM usia SMA dan sederajat adalah sebagai berikut:

$$\hat{y} = 0,073x_3 - 1,093(x_3 - 304,258)_+^{\frac{1}{4}} + 0,990(x_3 - 317,587)_+^{\frac{1}{4}} + 2,288(x_3 - 384,234)_+^{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} 0,073x_3 & x_3 < 304,258 \\ -1,02x_3 + 579,003 & 304,258 \leq x_3 < 317,587 \\ -0,03x_3 + 264,592 & 317,587 \leq x_3 < 384,234 \\ 2,258x_3 - 614,536 & x_3 \geq 384,234 \end{cases}$$

Jika terjadi kenaikan rasio murid-sekolah sebesar satu satuan pada kondisi dimana rasio murid-sekolah kurang dari 304,258, maka APM akan meningkat sebesar 0,07%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayapura, Nabire, Mimika, Boven Digoel, Yahukimo, Pegunungan Bintang, Tolikara, Keerom, Waropen, Supiori, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, dan Deiyai.

Jika terjadi kenaikan rasio murid-sekolah sebesar satu satuan pada kondisi dimana rasio murid-sekolah berada dalam interval 304,26 hingga 317,59, maka APM akan menurun sebesar 1,02%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayawijaya.

Jika terjadi kenaikan rasio murid-sekolah sebesar satu satuan pada kondisi dimana rasio murid-sekolah berada dalam interval 317,59 hingga 384,23, maka APM akan menurun sebesar 0,03%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Biak Numfor, Puncak Jaya, Mappi, dan Sarmi.

Jika terjadi kenaikan rasio murid-sekolah sebesar satu satuan pada kondisi dimana rasio murid-sekolah lebih dari sama dengan 384,234, maka APM akan meningkat sebesar 2,25%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Kepulauan Yapen, Painai, Asmat, dan Kota Jayapura.

Apabila variabel  $X_1, X_2, X_3,$  dan  $X_5$  diasumsikan konstan, maka pengaruh variabel  $X_4$  atau angka melek huruf terhadap APM usia SMA dan sederajat adalah sebagai berikut:

$$\hat{y} = 0,480x_4 - 12,343(x_4 - 84,093)_+^{\frac{1}{4}} + 12,280(x_4 - 85,860)_+^{\frac{1}{4}} + 1,222(x_4 - 94,698)_+^{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} 0,480x_4 & x_4 < 84,093 \\ -11,863x_4 + 1037,96 & 84,093 \leq x_4 < 85,860 \\ 0,417x_4 - 16,401 & 85,860 \leq x_4 < 94,698 \\ 1,639x_4 - 132,122 & x_4 \geq 94,698 \end{cases}$$

Jika terjadi kenaikan angka melek huruf sebesar 1% pada kondisi dimana angka melek huruf kurang dari 84,09%, maka APM akan meningkat sebesar 0,48%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayawijaya, Painai, Puncak Jaya, Yahukimo, Pegunungan Bintang, Tolikara, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, dan Deiyai.

Jika terjadi kenaikan angka melek huruf sebesar 1% pada kondisi dimana angka melek huruf berada dalam interval 84,09% hingga 85,86%, maka APM akan menurun sebesar 11,86%.

Jika terjadi kenaikan angka melek huruf sebesar 1% pada

kondisi dimana angka melek huruf berada dalam interval 85,86% hingga 94,7%, maka APM akan meningkat sebesar 0,42%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Kepulauan Yapen, Mappi, Asmat, dan Mamberamo Raya.

Jika terjadi kenaikan angka melek huruf sebesar 1% pada kondisi dimana angka melek huruf lebih dari sama dengan 94,7%, maka APM akan meningkat sebesar 1,64%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Jayapura, Nabire, Biak Numfor, Mimika, Boven Digoel, Sarmi, Keerom, Waropen, Supiori, dan Kota Jayapura.

Apabila variabel  $X_1, X_2, X_3,$  dan  $X_4$  diasumsikan konstan, maka pengaruh variabel  $X_5$  atau harapan lama sekolah terhadap APM usia SMA dan sederajat adalah sebagai berikut:

$$\hat{y} = 1,551 x_5 + 62,457 (x_5 - 11,236) \frac{1}{4} - 58,184 (x_5 - 11,654) \frac{1}{4} - 32,755 (x_5 - 13,745) \frac{1}{4}$$

$$\hat{y} = \begin{cases} 1,551 x_5 & x_5 < 11,236 \\ 64,008 x_5 - 701,767 & 11,236 \leq x_5 < 11,654 \\ 7,824 x_5 - 46,999 & 11,654 \leq x_5 < 13,745 \\ -24,931 x_5 + 403,219 & x_5 \geq 13,745 \end{cases}$$

Jika terjadi kenaikan harapan lama sekolah sebesar satu tahun pada kondisi dimana harapan lama sekolah kurang dari 11,24 tahun, maka APM akan meningkat sebesar 1,55%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Painai, Puncak Jaya, Boven Digoel, Mappi, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, dan Deiyai.

Jika terjadi kenaikan harapan lama sekolah sebesar satu tahun pada kondisi dimana harapan lama sekolah berada dalam interval 11,24 tahun hingga 11,65 tahun, maka APM akan meningkat sebesar 64,01%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Nabire.

Jika terjadi kenaikan harapan lama sekolah sebesar satu tahun pada kondisi dimana harapan lama sekolah berada dalam interval 11,65 tahun hingga 13,75 tahun, maka APM akan meningkat sebesar 7,82%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Jayawijaya, Kepulauan Yapen, Mimika, Sarmi, Keerom, Waropen, Supiori, dan Mamberamo Raya.

Jika terjadi kenaikan harapan lama sekolah sebesar satu tahun pada kondisi dimana harapan lama sekolah lebih dari sama dengan 13,75 tahun, maka APM akan menurun sebesar 24,93%. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayapura, Biak Numfor, dan Kota Jayapura.

### C. Pengujian Asumsi Residual

Pengujian asumsi residual dilakukan untuk mengetahui apakah residual data memenuhi asumsi identik, independen, dan berdistribusi normal.

#### 1) Asumsi Residual Identik

Hasil pengujian asumsi residual identik menggunakan Uji Glejser adalah sebagai berikut.

Hipotesis:

$H_0$  : residual identik (tidak terdapat heteroskedastisitas)

$H_1$  : residual tidak identik (terdapat heteroskedastisitas)

Tabel 8 menunjukkan bahwa hasil Uji Glejser adalah gagal tolak  $H_0$  karena nilai F sebesar 1,225 kurang dari  $F_{0,05(20;8)}$

sebesar 3,150 dan diperkuat oleh  $p$ -value sebesar 0,403 lebih dari taraf signifikan sebesar 0,05. Artinya, residual data memenuhi asumsi identik atau tidak terdapat heteroskedastisitas.

#### 2) Asumsi Residual Independen

Hasil pengujian asumsi residual independen menggunakan Uji Durbin Watson adalah sebagai berikut.

Hipotesis:

$H_0$  : residual independen (tidak terdapat autokorelasi)

$H_1$  : residual tidak independen (terdapat autokorelasi)

Tabel 9 menunjukkan bahwa hasil Uji Durbin Watson adalah gagal tolak  $H_0$  karena memiliki  $p$ -value sebesar 0,64 lebih dari taraf signifikan sebesar 0,05. Tetapi, jika dilihat dari daerah pengujiannya maka hasilnya tidak dapat disimpulkan karena nilai  $d$  sebesar 2,561 lebih dari  $4-d_U$  dan kurang dari  $4-d_L$ .

#### 3) Asumsi Residual Berdistribusi Normal

Hasil pengujian asumsi residual berdistribusi normal menggunakan Uji Kolmogorov-Smirnov adalah sebagai berikut:

Hipotesis:

$H_0$  : residual identik (tidak terdapat heteroskedastisitas)

$H_1$  : residual tidak identik (terdapat heteroskedastisitas)

Tabel 10 menunjukkan bahwa hasil Uji Kolmogorov-Smirnov adalah gagal tolak  $H_0$  karena nilai KS sebesar 0,124 kurang dari  $KS_{(0,05;29)}$  sebesar 0,246 dan diperkuat oleh  $p$ -value sebesar 0,722 lebih dari taraf signifikan sebesar 0,05. Artinya, residual data memenuhi asumsi distribusi normal.

## V. KESIMPULAN

Kesimpulan yang didapat berdasarkan hasil analisis adalah persentase penduduk miskin, rasio murid-guru, rasio murid-sekolah, angka melek huruf, dan harapan lama sekolah berpengaruh signifikan terhadap Angka Partisipasi Murni (APM) usia SMA dan sederajat di Provinsi Papua tahun 2019, sehingga model ditulis sebagai berikut:

$$\hat{y} = -37,516 + 0,087x_1 + 89,205 (x_1 - 32,946) \frac{1}{4} - 114,839 (x_1 - 34,136) + 41,335(x_1 - 40,082) \frac{1}{4} + 0,617 x_2 - 165,631(x_2 - 17,908) \frac{1}{4} + 190,940 (x_2 - 18,627) \frac{1}{4} + 7,866(x_2 - 22,222) \frac{1}{4} + 0,073 x_3 - 1,093 (x_3 - 304,258) \frac{1}{4} + 0,990 (x_3 - 317,587) \frac{1}{4} + 2,288(x_3 - 384,234) \frac{1}{4} + 0,480 x_4 - 12,343 (x_4 - 84,093) \frac{1}{4} + 12,280 (x_4 - 85,860) \frac{1}{4} + 1,222 (x_4 - 94,698) \frac{1}{4} + 1,551 x_5 + 62,457 (x_5 - 11,236) \frac{1}{4} - 58,184 (x_5 - 11,654) \frac{1}{4} - 32,755 (x_5 - 13,745) \frac{1}{4}$$

Koefisien determinasi ( $R^2$ ) yang diperoleh dari model tersebut sebesar 98,84% dengan MSE sebesar 16,786 serta asumsi residual yang telah terpenuhi, sehingga pemodelan yang dilakukan dapat dikatakan baik.

## DAFTAR PUSTAKA

[1] I. N. Budiantara, "Spline dalam Regresi Nonparametrik: Sebuah Pemodelan Statistik Masa Kini dan Masa Mendatang," Departemen

- Statistika: Institut Teknologi Sepuluh Nopember, 2009.
- [2] W. N. Daniel, *Applied Nonparametric Statistics*, 2nd ed. Boston: Houghton Mifflin, 1999, ISBN: 978-0534381943.
- [3] L. P. S. Pratiwi, "Pemilihan titik knot optimal model spline truncated dalam regresi nonparametrik multivariabel dengan GCV," *J. Mat.*, vol. 10, no. 2, pp. 78--90, 2020.
- [4] N. R. Draper and H. Smith, *Applied Regression Analys*. New Jersey: John Wiley & Sons, 1998, ISBN: 978-0471170822.
- [5] D. N. Gujarati, *Basic Econometric Analysis*. New Jersey: McGraw-Hill, 2004, ISBN: 978-0073375779.
- [6] I. Ghazali, *Aplikasi Analisis Multivariate dengan Program SPSS*. Semarang: Badan Penebit Universtas Diponogoro, 2006, ISBN: 9789797043001.