

# Pemodelan Faktor yang Mempengaruhi Persentase Anak Putus Sekolah di Jawa Timur Menggunakan Regresi Nonparametrik *Spline Truncated*

Delila Ramadanti Bidari dan I Nyoman Budiantara  
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)  
*e-mail: i\_nyoman\_b@statistika.its.ac.id*

**Abstrak**—Dunia pendidikan Indonesia telah memberlakukan wajib belajar 12 tahun, sebagai upaya pemerintah dalam meningkatkan kesejahteraan dan kemajuan bangsa melalui bidang pendidikan. Namun permasalahan di dunia pendidikan masih sering ditemui, salah satunya adalah anak putus sekolah. Meski peningkatan pendidikan sudah dilakukan oleh pemerintah, namun permasalahan putus sekolah masih banyak terjadi di Indonesia, khususnya di Provinsi Jawa Timur. Penyebab permasalahan putus sekolah dipengaruhi beberapa faktor. Untuk mengetahui faktor yang mempengaruhi dilakukan penelitian menggunakan regresi nonparametrik *spline truncated*. Karena pola hubungan yang ditunjukkan antar anak putus sekolah dengan faktor-faktor yang diduga mempengaruhinya tidak mengikuti pola tertentu. Berdasarkan nilai GCV yang paling minimum, model terbaik adalah menggunakan kombinasi titik knot (2, 3, 3, 2, 1). Hasil pengujian signifikansi parameter menunjukkan bahwa seluruh variabel yang digunakan dalam penelitian berpengaruh signifikan terhadap persentase anak putus sekolah di Jawa Timur. Variabel yang digunakan adalah persentase penduduk miskin, besaran upah minimum rata-rata, angka melek huruf, rasio guru terhadap murid, dan rata-rata jumlah anggota keluarga. Hasil pengujian asumsi residual menunjukkan semua asumsi terpenuhi dengan nilai koefisien determinasi dari model ini sama dengan 77,67%.

**Kata Kunci**—Anak Putus Sekolah, GCV, Jawa Timur, Regresi Nonparametrik, *Spline Truncated*.

## I. PENDAHULUAN

**P**EMBANGUNAN sebuah negara dapat dipengaruhi oleh banyak hal salah satunya pendidikan, dimana pendidikan memiliki peranan penting untuk mengembangkan Sumber Daya Manusia (SDM) yang dimiliki oleh suatu negara, dalam hal ini pendidikan membawa individu menuju kehidupan yang lebih baik. Undang-undang No. 39 Tahun 1999 Tentang Hak Asasi Manusia Pasal 60 Ayat (1) menyatakan bahwa “setiap anak berhak memperoleh pendidikan dan pengajaran dalam rangka pengembangan pribadi sesuai dengan minat, bakat dan tingkat kecerdasannya”. Hal tersebut berarti bahwa setiap anak Indonesia berhak memperoleh pendidikan, oleh karena itu pemerintah melakukan berbagai upaya dalam usaha meningkatkan kualitas pendidikan di Indonesia agar dapat lebih merata di seluruh wilayah namun saat ini masih banyak masalah dalam bidang pendidikan di Indonesia, salah satunya adalah anak putus sekolah, seorang anak dikatakan putus sekolah ketika anak tersebut tidak mampu menyelesaikan suatu jenjang pendidikan, sehingga tidak dapat melanjutkan *study* ke jenjang pendidikan berikutnya [1]. UNESCO menyatakan bahwa jumlah anak yang dirampas hak-hak pendidikan di seluruh dunia mencapai 263

juta anak, termasuk 142 juta orang di jenjang sekolah tingkat atas. Tahun 2016 secara nasional Jawa Timur menempati posisi tertinggi kedua setelah Jawa Barat dengan jumlah anak putus sekolah 21.023 siswa [2].

Jumlah anak putus sekolah jenjang pendidikan sekolah dasar sebanyak 39.164, jenjang sekolah menengah pertama 38.706, jenjang sekolah menengah atas 36.419 dan jenjang sekolah menengah kejuruan sebanyak 72.744. Indonesia menempati urutan ke 2 di dunia dengan angka putus sekolah jenjang SMA tertinggi yang mencapai 60% dari masyarakat kelompok usia 25-34 tahun yang gagal menyelesaikan pendidikan menengah [3]. Terdapat beberapa faktor yang mempengaruhi jumlah anak putus sekolah antara lain faktor ekonomi keluarga dan pendidikan orang tua yang masih menjadi kendala bagi masyarakat untuk memberikan hak pendidikan anak wajib belajar yang telah ditentukan oleh pemerintah.

Selain itu faktor yang mempengaruhi anak putus sekolah terbagi menjadi dua yaitu faktor internal dan faktor eksternal. Faktor internal terdiri dari fasilitas pendidikan yang tersedia hingga metode pendidikan yang diterapkan, sedangkan faktor eksternal antara lain faktor ekonomi dan faktor sosial, dimana faktor eksternal tersebut dapat ditinjau dari pendidikan orang tua, keterbukaan masyarakat terhadap pengetahuan, kemiskinan, beban tanggungan keluarga, lokasi tinggal yang jauh dari kota hingga suku dan status sosial [4].

Berdasarkan referensi dan penelitian sebelumnya dengan kasus yang serupa dan metode yang berbeda, penelitian ini akan memodelkan faktor-faktor yang mempengaruhi persentase anak putus sekolah di Jawa Timur sebagai provinsi dengan jumlah Anak putus sekolah terbanyak nomor 2 di Indonesia pada tahun ajaran 2017/2018 dan 2018/2019 menggunakan regresi nonparametrik *spline truncated*. Hal ini dikarenakan kurva regresi yang dibentuk antara variabel prediktor dengan variabel respon pada data anak putus sekolah di Jawa Timur tahun 2017, tidak mengikuti pola tertentu. Pemilihan metode regresi nonparametrik *spline truncated* didasari oleh kelebihan dari metode ini dibandingkan metode regresi nonparametrik *spline* lainnya, yaitu metode lebih sederhana, interpretasi mudah, dan perumusan matematis lebih sederhana [5]. Faktor-faktor yang diduga mempengaruhi persentase anak putus sekolah di Jawa Timur diantaranya adalah persentase penduduk miskin, besaran upah minimum rata-rata, angka melek huruf, rasio murid-guru dan rata-rata jumlah anggota keluarga. Diharapkan dengan hasil penelitian ini dapat menjadi bahan pertimbangan atau rekomendasi bagi pemerintah Indonesia

Tabel 1.  
Analisis ragam (ANOVA) uji parameter

Sumber Variasi	Df	Sum of Square (SS)	Mean Square (MS)	$F_{hitung}$
Regresi	$p + r$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{SS_{regresi}}{df_{regresi}}$	$\frac{MS_{regresi}}{MS_{error}}$
Error	$n - (p + r) - 1$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\frac{SS_{error}}{df_{error}}$	
Total	$n - 1$	$\sum_{i=1}^n (y_i - \bar{y})^2$		

Tabel 2.  
Kebutuhan data penelitian

No.	Variabel	Keterangan
1	Y	Persentase Kasus Anak Putus Sekolah
2	$X_1$	Persentase Penduduk Miskin
3	$X_2$	Besaran Upah Minimum Rata-Rata
4	$X_3$	Angka Melek Huruf
5	$X_4$	Rasio Guru Terhadap Murid
6	$X_5$	Rata-Rata Jumlah Anggota Keluarga

kususnya pemerintah daerah Jawa Timur dalam rangka mengurangi anak putus sekolah.

## II. TINJAUAN PUSTAKA

### A. Regresi Nonparametrik Spline Truncated

Regresi nonparametrik *spline truncated* merupakan metode yang paling banyak digunakan pada regresi nonparametrik. Bentuk kurva *spline* terpotong-potong sehingga mampu mengatasi perubahan pola data pada sub interval tertentu. Pada metode regresi nonparametrik *spline truncated* digunakan bantuan titik-titik knot. Titik knot merupakan titik dimana terjadi pola perubahan perilaku dari suatu fungsi pada selang yang berbeda [6]. Fungsi *spline truncated* multivariabel berorde  $q$  dengan titik knot  $K_1, K_2, \dots, K_r$ . Dapat dituliskan menjadi persamaan sebagai berikut:

$$\sum_{j=1}^p f_j(x_{ji}) = \sum_{j=1}^p \sum_{u=0}^q \beta_{ju} x_{ji}^u + \sum_{j=1}^p \sum_{k=1}^r \beta_{j(q+k)} (x_{ji} - K_{jk})_+^q \quad (1)$$

Model regresi nonparametrik *spline truncated* sebagai berikut,

$$y_i = \sum_{j=1}^p \sum_{u=0}^q \beta_{ju} x_{ji}^u + \sum_{j=1}^p \sum_{k=1}^r \beta_{j(q+k)} (x_{ji} - K_{jk})_+^q + \varepsilon_i \quad (2)$$

Fungsi *truncated*  $(x_{ji} - K_{jk})_+^q$  akan menghasilkan persamaan berikut:

$$(x_{ji} - K_{jk})_+^q = \begin{cases} (x_{ji} - K_{jk})^q, & x_{ji} > K_{jk} \\ 0 & , x_{ji} \leq K_{jk} \end{cases} \quad (3)$$

Fungsi *spline* yang digunakan adalah *Spline* linear dengan derajat  $q = 1$ .

### B. Pemilihan Titik Knot Optimum

Model regresi *spline* terbaik merupakan model yang memiliki titik knot optimal. Titik knot merupakan titik yang terdapat pada perubahan pola perilaku fungsi. Salah satu metode yang biasa digunakan untuk memilih titik knot optimal adalah metode *Generalized Cross Validation* (GCV). Menggunakan metode GCV didasari oleh kelebihan yang dimiliki metode ini, yaitu mempunyai sifat optimal asimtotik, tidak memuat varians populasi ( $\sigma^2$ ) yang tidak diketahui, *invariance* terhadap transformasi. Model regresi *spline* terbaik diperoleh dari titik knot optimal dengan melihat nilai GCV terkecil [7]. Metode GCV dapat dituliskan sebagai berikut.

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{[n^{-1} \text{trace}(I-A)]^2} \quad (4)$$

dimana  $I$  merupakan matriks identitas,  $n$  adalah jumlah pengamatan,  $K = (K_1, K_2, \dots, K_r)$  merupakan titik-titik knot,

$$MSE(K_1, K_2, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (5)$$

serta  $A = X(X^T X)^{-1} X^T$ .

### C. Pengujian Parameter Model Regresi

Pengujian parameter model dilakukan untuk menentukan variabel prediktor yang berpengaruh terhadap variabel respon. Terdapat dua tahap pengujian parameter yaitu pengujian secara serentak dan secara parsial.

#### 1) Uji Serentak

Hipotesis:

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

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

Dimana  $p$  adalah jumlah variabel prediktor dan  $r$  adalah jumlah titik knot. Statistik uji dalam pengujian serentak menggunakan uji  $F$  seperti pada persamaan berikut ini.

$$F_{hitung} = \frac{MS_{regresi}}{MS_{residual}} \quad (6)$$

Pengujian parameter model secara serentak dapat disajikan menggunakan *Analysis of Varians* (ANOVA) yang disajikan dalam Tabel 1. Nilai  $p + r$  pada Tabel 1 merupakan banyak parameter dalam model regresi nonparametrik *spline* kecuali  $\beta_0$ . Tolak  $H_0$  jika  $F_{hitung} > F_{(\alpha; (p+r); n-(p+r)-1)}$ , sehingga menghasilkan kesimpulan bahwa minimal ada satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon [8].

#### 2) Uji Parsial

Pengujian secara parsial bertujuan untuk mengetahui parameter mana yang berpengaruh dan tidak berpengaruh signifikan terhadap model regresi.

Hipotesis:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p + r$$

Pengujian secara parsial dilakukan dengan menggunakan uji  $t$  [8]. Statistik uji yang digunakan adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (7)$$

$SE(\hat{\beta}_j)$  adalah *standart error*  $\hat{\beta}_j$  dimana  $SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)}$  dengan  $\text{Var}(\hat{\beta}_j)$  merupakan elemen diagonal utama ke- $j$  dari matriks  $\text{Var}(\hat{\beta})$ ,  $j = 1, 2, \dots, p + r$ . Tolak  $H_0$  jika  $t > t_{\frac{\alpha}{2}; (n-(p+r)-1)}$  dan  $t < -t_{\frac{\alpha}{2}; (n-(p+r)-1)}$ , sehingga menghasilkan kesimpulan bahwa variabel prediktor ke- $n$  berpengaruh signifikan terhadap variabel respon [8].

### D. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi adalah kuantitas yang dapat menjelaskan sumbangan variabel prediktor terhadap variabel

Tabel 3. Karakteristik persentase kasus anak putus sekolah dan faktor-faktor yang diduga berpengaruh

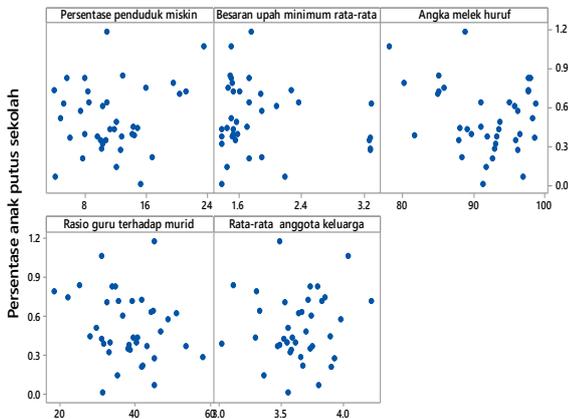
Variabel	Mean	Varians	Min	Maks
Y	0,511	0,069	0,011	1,9803
X <sub>1</sub>	11,63	22,28	4,17	23,56
X <sub>2</sub>	1,872	0,372	1,389	3,296
X <sub>3</sub>	91,806	29,494	78,120	98,740
X <sub>4</sub>	38,41	70,22	18,78	58,37
X <sub>5</sub>	3,643	0,06	3,03	4,23

Tabel 4. Perbandingan nilai GCV minimum tiap titik knot

No.	Knot	GCV Minimum
1	Satu Titik Knot	0,081545
2	Dua Titik Knot	0,0589843
3	Tiga Titik Knot	0,0594583
4	Kombinasi Knot (2, 3, 3, 2, 1)	0,044428

Tabel 5. Analysis of variance

Sumber Variasi	df	SS	MS	F	P-Value
Regresi	16	1,96832	0,12302	4,56541	0,000746
Error	21	0,56587	0,02695		
Total	37	2,53419			



Gambar 1. Scatterplot persentase anak putus sekolah di Jawa Timur tahun 2017 dan faktor-faktor yang diduga mempengaruhinya.

$$R^2 = \frac{SS_{Regresi}}{SS_{Total}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{8}$$

E. Pengujian Asumsi Residual

Residual yang dihasilkan harus memenuhi asumsi. Terdapat tiga asumsi yang harus dipenuhi yaitu identik, independen, dan berdistribusi normal.

1) Asumsi Identik

Asumsi identik biasa juga disebut homoskedastisitas yang berarti bahwa varians pada residual adalah identik. Kebalikannya adalah kasus heteroskedastisitas, yaitu jika kondisi varians residual tidak identik [8].

$$var(y_i) = var(\epsilon_i) = \sigma^2 ; i = 1, 2, \dots, n \tag{9}$$

Uji identik dapat menggunakan uji Glejser.

Hipotesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, n$$

Statistik uji yang digunakan adalah:

$$F_{hitung} = \frac{\sum_{i=1}^n (|\hat{\epsilon}_i| - |\epsilon_i|)^2}{\frac{p+r}{n-(p+r)-1}} \tag{10}$$

Tolak H<sub>0</sub> jika F<sub>hitung</sub> > F<sub>α;(v,n-v-1)</sub> atau p<sub>value</sub> < α yang berarti bahwa tidak terindikasi terdapat kasus homoskedastisitas.

2) Asumsi Independen

Merupakan asumsi dari model regresi yang mengharuskan tidak terdapat korelasi antar residual. Pendeteksian autokorelasi dapat dilakukan dengan membuat plot Autocorrelation Function (ACF). Berikut ini merupakan rumus untuk menghitung ACF.

respon. Semakin tinggi nilai R<sup>2</sup> yang dihasilkan suatu model, maka semakin baik pula variabel-variabel prediktor dalam model tersebut dalam menjelaskan variabilitas variabel respon. Berikut ini adalah rumus untuk mendapatkan nilai R<sup>2</sup> [9].

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (e_t - \bar{e}_t)(e_{t+k} - \bar{e}_t)}{\sum_{t=1}^n (e_t - \bar{e}_t)^2} \tag{11}$$

Apabila nilai  $\hat{\rho}_k$  keluar dari batas signifikan maka dapat dikatakan terdapat kasus autokorelasi. Sebaliknya jika tidak terdapat nilai  $\hat{\rho}_k$  yang keluar dari batas signifikan maka tidak terdapat kasus autokorelasi.

3) Asumsi Normalitas Kolmogorov Smirnov

Uji Kolmogorov-Smirnov dilakukan untuk mengetahui apakah suatu data telah mengikuti suatu distribusi normal [8].

Hipotesis:

$$H_0 : F_n(\epsilon) = F_0(\epsilon)$$

$$H_1 : F_n(\epsilon) \neq F_0(\epsilon)$$

Atau

$$H_0 : \text{residual berdistribusi normal}$$

$$H_1 : \text{residual tidak berdistribusi normal}$$

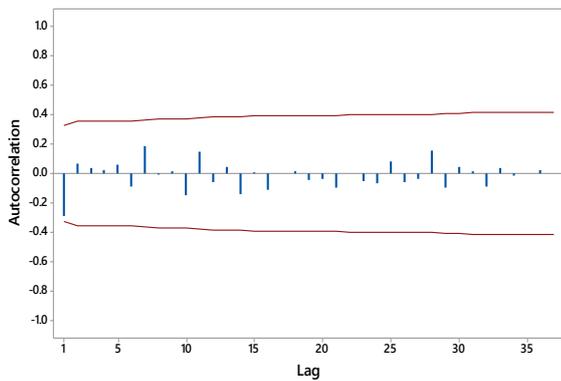
Statistik uji:

$$D = \text{maksimal} |F_n(\epsilon) - F_0(\epsilon)| \tag{12}$$

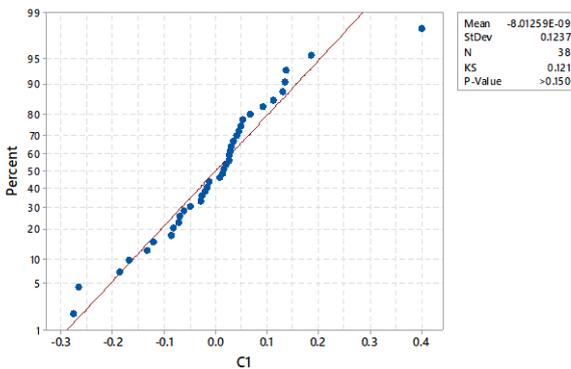
Tolak H<sub>0</sub> apabila |D| > D<sub>(1-α)</sub>. D<sub>(1-α)</sub> adalah nilai kritis untuk uji Kolmogorov Smirnov satu sampel, diperoleh dari tabel Kolmogorov Smirnov satu sampel, F<sub>n</sub>(ε) merupakan nilai peluang kumulatif (fungsi distribusi kumulatif) berdasarkan data sampel, F<sub>0</sub>(ε) adalah nilai peluang kumulatif (fungsi distribusi kumulatif) dibawah H<sub>0</sub>.

F. Putus Sekolah

Putus sekolah merupakan predikat yang diberikan kepada mantan peserta didik yang tidak mampu menyelesaikan suatu jenjang pendidikan, sehingga tidak dapat melanjutkan studinya ke jenjang pendidikan berikutnya. Misalnya seorang warga masyarakat atau anak yang hanya mengikuti pendidikan di SD sampai kelas lima, disebut sebagai putus sekolah SD. Demikian juga seorang warga masyarakat yang memiliki ijazah SD kemudian mengikuti pendidikan di SMP sampai kelas dua saja, disebut putus SMP, dan seterusnya. Sedangkan angka putus sekolah merupakan proporsi anak menurut kelompok usia yang sudah tidak bersekolah lagi atau tidak menamatkan suatu jenjang pendidikan tertentu. Adapun kelompok umur yang dimaksud adalah kelompok umur 7-12 tahun, 13-15 tahun dan 16-18 tahun [1].



Gambar 2. Plot ACF residual.



Gambar 3. Hasil uji kolmogrov smirnov dan plot normalitas residual.

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Dinas Pendidikan dan Badan Pusat Statistik (BPS) Jawa Timur. Data tersebut mengenai kasus Anak Putus Sekolah menurut kabupaten/kota di Jawa Timur tahun ajaran 2016/2017 dan 2017/2018. Unit observasi yang digunakan dalam penelitian ini merupakan 38 kabupaten/kota di Jawa Timur.

#### B. Variabel Penelitian

Variabel respon yang digunakan dalam penelitian dijelaskan pada Tabel 2.

#### C. Langkah Penelitian

Berikut adalah langkah-langkah analisis yang digunakan dalam melakukan penelitian.

1. Mendeskripsikan karakteristik kasus anak putus sekolah setiap kabupaten/kota di Jawa Timur tahun 2017 dan faktor-faktor yang diduga mempengaruhinya
2. Mengidentifikasi pola hubungan antara variabel respon dengan masing-masing variabel prediktor dengan menggunakan *scatterplot*
3. Menganalisis model kasus anak putus sekolah di Jawa Timur tahun 2017 menggunakan Regresi Nonparametrik *Spline Truncated* dengan langkah sebagai berikut:

Tabel 6. Hasil pengujian signifikansi parameter secara individu

Variabel	Parameter	Estimator	t	P-value
$x_1$	$\beta_0$	-4,153040	-1,287985	0,211763
	$\beta_1$	0,098357	3,358791	0,002972
	$\beta_2$	-0,144119	-3,905419	0,000814
$x_2$	$\beta_3$	0,514203	4,147620	0,000456
	$\beta_4$	0,773625	3,1015066	0,005404
	$\beta_5$	-1,191167	-2,010783	0,057363
$x_3$	$\beta_6$	1,412302	0,421119	0,677947
	$\beta_7$	0,929606	0,416012	0,681622
	$\beta_8$	0,079502	1,788966	0,088056
$x_4$	$\beta_9$	-0,088395	-2,014702	0,056919
	$\beta_0$	2,129819	5,160866	0,000041
	$\beta_1$	-2,702898	-4,326981	0,000297
$x_5$	$\beta_2$	-0,079565	-4,462191	0,000215
	$\beta_3$	0,091062	4,405023	0,000247
	$\beta_4$	-0,057291	-1,825821	0,082132
$x_6$	$\beta_5$	-0,506071	-2,151609	0,043218
	$\beta_6$	0,929133	1,305173	0,205955

Tabel 7. Hasil pengujian statistik uji glejser

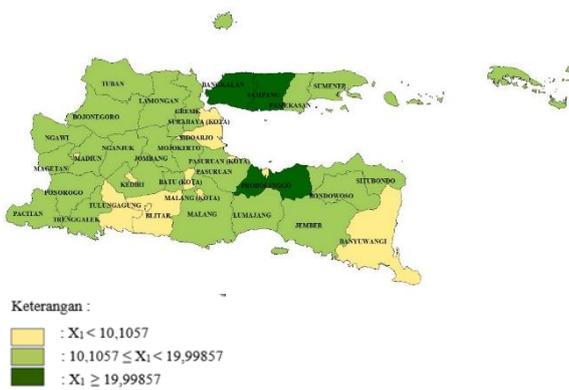
Sumber	Df	SS	MS	Fhit	P-value
Regresi	16	0,078798	0,004925	0,536256	0,896350
Error	21	0,192859	0,009184		
Total	37	0,271657			

- a. Memodelkan variabel respon menggunakan model regresi nonparametrik *spline truncated* dengan satu, dua, tiga dan kombinasi knot.
- b. Menentukan titik knot optimal berdasarkan nilai GCV yang paling minimum. Memperoleh model terbaik dengan titik knot optimal.
- c. Mendapatkan estimator parameter model.
- d. Melakukan pengujian signifikansi parameter secara serentak dan parsial.
- e. Melakukan uji asumsi residual identik, independen, dan berdistribusi normal (IIDN).
- f. Menghitung nilai  $R^2$
- g. Membuat interpretasi model dan menarik kesimpulan.

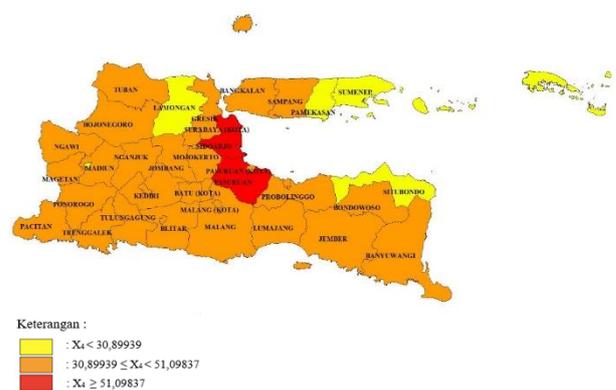
### IV. ANALISIS DAN PEMBAHASAN

#### A. Karakteristik Persentase Anak Putus Sekolah di Jawa Timur dan Faktor-Faktor yang Mempengaruhinya.

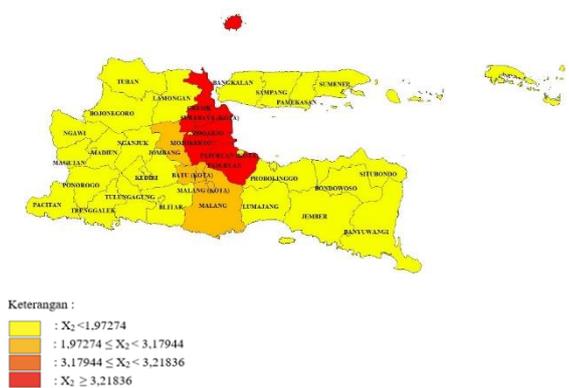
Karakteristik persentase anak putus sekolah di Jawa Timur beserta faktor-faktor yang diduga mempengaruhinya dapat diketahui melalui statistika deskriptif. Pada Tabel 3, diberikan informasi mengenai statistika deskriptif yang meliputi rata-rata, varians, nilai minimum serta maksimum. Kabupaten/kota dengan persentase anak putus sekolah tertinggi adalah Jember sama dengan 1,98%. Artinya dari 100 anak usia sekolah 2 orang diantaranya tidak dapat menyelesaikan atau melanjutkan jenjang pendidikan. Sedangkan kabupaten/kota dengan persentase anak putus sekolah terendah adalah Pacitan sama dengan 0,01%, artinya dari 100 anak usia sekolah tidak ada siswa yang mengalami putus sekolah.



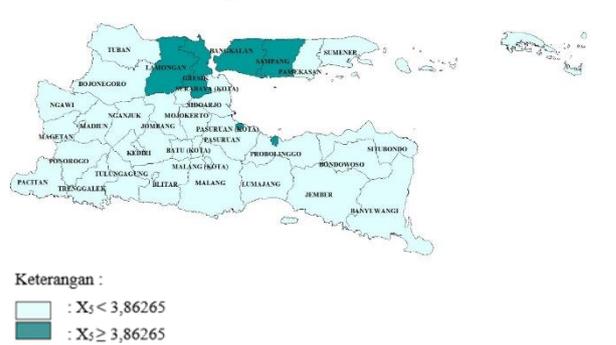
Gambar 4. Peta persebaran berdasarkan persentase penduduk miskin.



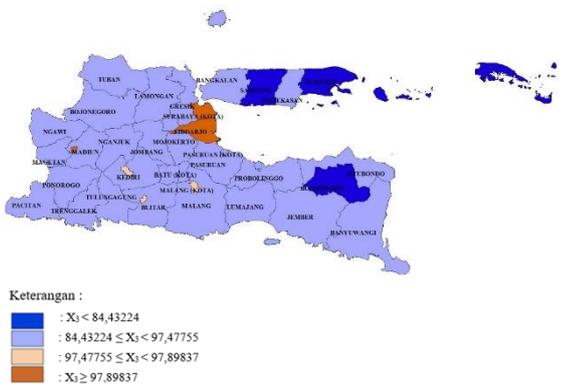
Gambar 7. Peta persebaran berdasarkan rasio guru terhadap murid.



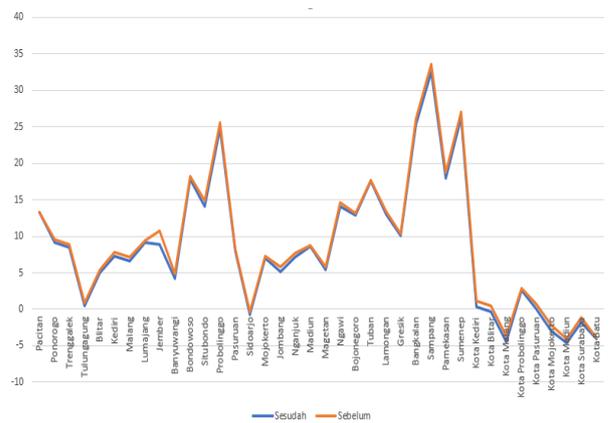
Gambar 5. Peta persebaran berdasarkan upah minimum rata-rata.



Gambar 8. Peta persebaran berdasarkan jumlah anggota keluarga.



Gambar 6. Peta persebaran berdasarkan angka melek huruf.



Gambar 9. Perbandingan persentase anak putus sekolah di Jawa timur sebelum dan sesudah pemodelan.

**B. Analisis Pola Hubungan Faktor-Faktor yang Diduga Mempengaruhi Persentase Anak Putus Sekolah di Jawa Timur**

Dengan menggunakan *scatterplot* dapat Mengidentifikasi pola hubungan antara variabel respon dengan variabel prediktor, maka dapat ditentukan komponen parametrik dan nonparametrik. Gambar 1 adalah *scatterplot* antara variabel respon dengan masing-masing variabel prediktor.

Gambar 1 menunjukkan bahwa hubungan antara persentase anak putus sekolah dengan lima faktor yang diduga mempengaruhinya tidak membentuk pola tertentu. Sehingga, variabel prediktor yang digunakan merupakan komponen nonparametrik dan cocok untuk dimodelkan dengan regresi nonparametrik *spline truncated*.

**C. Pemilihan Titik Knot Optimum**

Berdasarkan pemodelan yang telah dilakukan menggunakan satu, dua, tiga dan kombinasi titik knot

selanjutnya pemilihan model terbaik, didasari pada titik knot optimal yang ditunjukkan oleh nilai GCV yang paling minimum. Berdasarkan pemodelan yang telah dilakukan dengan berbagai titik knot (satu, dua, tiga, dan kombinasi), masing-masing nilai GCV pada tiap titik knot yang paling optimal tersebut dibandingkan. Berikut nilai GCV minimum pada tiap pemodelan menggunakan satu, dua, tiga dan kombinasi knot.

Tabel 4 menunjukkan nilai GCV yang paling minimum pada tiap titik knot, kombinasi titik knot pada titik (2, 3, 3, 2, 1) memiliki GCV paling minimum sehingga menghasilkan nilai yang paling optimal. Dengan dua titik knot pada variabel persentase penduduk miskin, tiga titik knot pada variabel besaran upah minimum rata-rata, tiga titik knot pada variabel angka melek huruf, dua titik knot pada variabel rasio guru terhadap murid, dan satu titik knot pada variabel rata-rata jumlah anggota keluarga.

#### D. Penaksiran Parameter

Berdasarkan nilai GCV yang paling minimum yaitu pada titik knot kombinasi (2, 3, 3, 2, 1) maka titik tersebut digunakan untuk memodelkan data persentase anak putus sekolah di Jawa Timur tahun 2017. Sehingga dapat diestimasi parameter untuk model regresi nonparametrik *spline truncated* dengan titik kombinasi (2, 3, 3, 2, 1). Estimasi parameter menggunakan metode *Ordinary Least Square* (OLS) sebagai berikut.

$$\begin{aligned} \hat{y}_i = & -4,15304 + 0,09836x_{i1} \\ & - 0,14412(x_{i1} - 10,1057)_+ \\ & + 0,514203(x_{i1} - 19,99857)_+ + \\ & 0,77363x_{i2} - 1,19117(x_{i2} - 1,97274)_+ + \\ & 1,41230(x_{i2} - 3,17944)_+ + \\ & 0,92961(x_{i2} - 3,21836)_+ + 0,0795x_{i3} - \\ & 0,08839(x_{i3} - 84,43224)_+ + \\ & 2,12982(x_{i3} - 97,47755)_+ - \\ & 2,70289(x_{i3} - 97,89837)_+ - 0,079576x_{i4} + \\ & 0,09106(x_{i4} - 30,89939)_+ - \\ & 0,05729(x_{i4} - 51,09837)_+ - \\ & 0,50607x_{i5} + 0,92913(x_{i5} - 3,86265)_+ \end{aligned}$$

#### E. Pengujian Signifikansi Parameter

Pengujian parameter model regresi nonparametrik *spline truncated* terdiri dari pengujian secara serentak dan pengujian secara parsial atau individu sebagai berikut.

##### 1) Uji Serentak

Hipotesis yang digunakan untuk pengujian parameter secara serentak adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{16} = 0$$

$$H_1 : \text{Minimal terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, 16.$$

Dengan menggunakan taraf kepercayaan 95% diperoleh  $F_{(0,05;16;21)} = 2,156263$ . Berikut hasil pengujian parameter secara serentak ditampilkan pada Tabel 5.

Berdasarkan Tabel 5 nilai statistik uji  $F_{hitung}$  sama dengan 4,565413. Karena nilai statistik uji  $F_{hitung} > F_{(0,05;16;21)}$ , maka menghasilkan keputusan Tolak  $H_0$ . Selain itu diperoleh  $p_{value} < 5\%$  dan diperoleh keputusan tolak  $H_0$ . Menunjukkan bahwa minimal terdapat satu parameter yang berpengaruh signifikan pada model, atau secara bersama terdapat minimal satu variabel prediktor berpengaruh signifikan terhadap persentase anak putus sekolah di Jawa Timur.

##### 2) Uji Parsial

Berikut hipotesis untuk melakukan uji parameter secara parsial.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j = 1, 2, \dots, 16.$$

Pengujian parameter secara parsial akan dibandingkan dengan  $t_{0,025;21} = 2,080$  dan  $-t_{0,025;21} = -2,080$  dengan menggunakan taraf kepercayaan 95%.

Berdasarkan Tabel 6 diperoleh tujuh nilai statistik uji dimana  $t > -t_{(0,025;17)}$  sehingga gagal tolak  $H_0$ , yaitu parameter  $\beta_5$  (variabel  $x_2$ ),  $\beta_6$  (variabel  $x_2$ ),  $\beta_7$  (variabel  $x_2$ ),  $\beta_8$  (variabel  $x_3$ ),  $\beta_9$  (variabel  $x_3$ ),  $\beta_{14}$  (variabel  $x_4$ ), dan  $\beta_{16}$  (variabel  $x_5$ ). Hal ini menunjukkan tujuh parameter tersebut tidak berpengaruh signifikan terhadap model. Sedangkan sembilan parameter lainnya berpengaruh signifikan terhadap model, karena memiliki nilai statistik uji  $t > t_{(0,025;17)}$  dan  $t < -t_{(0,025;17)}$ . Meski terdapat beberapa parameter yang

tidak signifikan, setiap variabel masih memiliki nilai parameter yang signifikan. Sehingga dikatakan seluruh variabel prediktor berpengaruh signifikan terhadap variabel respon secara parsial.

#### F. Pengujian Asumsi Residual

Tiga asumsi residual harus dipenuhi sebuah model dalam analisis regresi nonparametrik *spline truncated*, yaitu asumsi residual identik, independen dan berdistribusi normal.

##### 1) Asumsi Identik

Hasil pengujian asumsi residual identik adalah sebagai berikut.

Hipotesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, 38$$

Dengan menggunakan taraf kepercayaan 95% diperoleh  $F_{(0,05;16;21)} = 2,156263$ . Hasil pengujian asumsi identik ditampilkan pada Tabel 7.

Berdasarkan Tabel 7 diperoleh nilai  $p_{value} > 5\%$  dengan  $p_{value} = 0,8963502$ , maka diperoleh keputusan gagal tolak  $H_0$ , selain itu nilai statistik uji  $F_{hitung}$  sama dengan 0,5362559. Karena nilai statistik uji  $F_{hitung} < F_{(0,05;16;21)}$ , maka diperoleh keputusan gagal tolak  $H_0$ . Dapat disimpulkan bahwa tidak terjadi heterokedastisitas pada model, sehingga asumsi residual identik terpenuhi.

##### 2) Asumsi Independen

Pengujian asumsi residual independen salah satu cara untuk mengetahui apakah terjadi kasus autokorelasi atau tidak adalah dengan melihat plot dari *Autocorrelation Function* (ACF). Berikut ini adalah plot dari *Autocorrelation Function* (ACF) residual. Berdasarkan Gambar 2 terlihat bahwa tidak terdapat autokorelasi yang melewati atau keluar dari interval konfiden. Sehingga dapat disimpulkan bahwa tidak terjadi kasus autokorelasi.

##### 3) Asumsi Normal

Hasil pengujian asumsi distribusi normal menggunakan uji *Kolmogorov Smirnov* yang ditunjukkan pada Gambar 3. Pada Gambar 3 titik plot residual menunjukkan penyebaran titik plot residual berada pada sumbu diagonal grafik, hal ini mengindikasikan residual berdistribusi normal. Selain itu berdasarkan nilai pengujian *Kolmogorov Smirnov* diperoleh  $D = 0,121$  dengan  $p_{value} > 0,150$ . Karena nilai statistik uji  $D < D_{(1-0,05)}$ , dan  $p_{value} > 5\%$  menghasilkan keputusan gagal tolak  $H_0$ . Maka ditarik kesimpulan bahwa residual telah mengikuti pola distribusi normal.

#### G. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi menunjukkan seberapa baiknya model yang digunakan. Dengan perhitungan  $R^2$  sebagai berikut.

$$\begin{aligned} R^2 &= \frac{SS_{Regresi}}{SS_{Total}} \times 100\% \\ &= 77,6\% \end{aligned}$$

Berdasarkan hasil perhitungan diperoleh nilai koefisien determinasi ( $R^2$ ) 77,67%. Sehingga dapat disimpulkan model regresi nonparametrik *spline truncated* dengan titik kombinasi (2,3,3,2,1) mampu menjelaskan variabilitas persentase anak putus sekolah di Jawa Timur sama dengan 77,67%. Dengan kata lain persentase anak putus sekolah di Jawa Timur dapat dijelaskan kelima variabel prediktor yang

digunakan dalam penelitian ini, sama dengan 77,67% sedangkan sisanya dijelaskan oleh variabel lain yang tidak terdapat dalam model.

*H. Interpretasi Model*

Pada pembahasan sebelumnya model telah memenuhi ketiga asumsi residual. Sehingga model regresi nonparametrik *spline truncated* menggunakan titik kombinasi 2,3,3,2,1 sudah layak digunakan untuk memodelkan persentase anak putus sekolah di Jawa Timur. Dengan interpretasi dari model sebagai berikut.

1) *Mengasumsikan Semua Variabel Selain Variabel  $X_1$*

Mengasumsikan semua variabel selain variabel  $X_1$  konstan, maka persamaan regresi dari persentase penduduk miskin ( $X_1$ ) terhadap persentase kasus anak putus sekolah ( $Y$ ) adalah sebagai berikut.

$$\hat{y} = \begin{cases} -4,15304 + 0,09836x_1 & ; x_1 < 10,1057 \\ -2,99661 - 0,04570x_1 & ; 10,106 \leq x_1 < 19,99 \\ -13,2799 + 0,4685x_1 & ; x_1 \geq 19,99857 \end{cases}$$

Interval  $x_1 < 10,1057$  menunjukkan apabila persentase penduduk miskin di suatu wilayah kurang dari 10,1057% maka setiap kenaikan persentase penduduk miskin sebesar 1% akan meningkatkan persentase anak putus sekolah sebesar 0,9836 kasus. Pada interval 10,1057% sampai 19,99857% maka setiap kenaikan persentase penduduk miskin sebesar 1% mengakibatkan persentase anak putus sekolah turun sebanyak 0,04570, interval  $x_1 \geq 8,84$  menunjukkan bahwa apabila persentase penduduk miskin suatu wilayah berada pada interval lebih dari 19,99857%, maka setiap kenaikan persentase penduduk miskin sebesar 1% mengakibatkan persentase anak putus sekolah akan naik 0,4685. Peta persebaran persentase penduduk miskin dapat dilihat pada Gambar 4.

2) *Mengasumsikan Semua Variabel Selain Variabel  $X_2$*

Berikut adalah persamaan persamaan regresi dari besaran upah minimum rata-rata ( $X_2$ ) terhadap persentase kasus anak putus sekolah di Jawa Timur ( $Y$ ).

$$\hat{y} = \begin{cases} -4,1530 + 0,7736x_2 & ; x_2 < 1,97274 \\ -1,8032 - 0,4175x_2 & ; 1,9727 \leq x_2 < 3,1794 \\ -6,2935 + 0,9948x_2 & ; 3,1794 \leq x_2 < 3,218 \\ -9,2853 + 1,9244x_2 & ; x_2 \geq 3,21836 \end{cases}$$

Pada interval  $x_2 < 1,97274$  menunjukkan apabila besaran upah minimum rata-rata disuatu provinsi kurang dari 1,97274 juta maka setiap kenaikan sebesar 1 juta akan mengakibatkan persentase anak putus sekolah di Jawa Timur naik sebesar 0,77363. Interval 1,97274 sampai 3,17944 maka setiap kenaikan sebesar 1 juta akan mengakibatkan persentase anak putus sekolah di Jawa Timur turun sebesar 0,41754. Wilayah yang tergolong dalam interval kedua adalah Jombang, Kota Batu, Kota Malang dan Malang. Pada interval 3,17944 sampai 3,21836 menunjukkan bahwa setiap kenaikan sebesar 1 juta akan mengakibatkan persentase anak putus sekolah di Jawa Timur naik sebesar 0,99476. Wilayah yang tergolong dalam interval tersebut adalah Mojokerto. Sedangkan pada interval lebih dari 3,21836 kenaikan sebesar 1 juta akan mengakibatkan persentase anak putus sekolah di Jawa Timur naik sebesar 1,92437. Provinsi yang tergolong dalam interval keempat ialah Pasuruan, Sidoarjo, Gresik dan Kota Surabaya.

Peta persebaran berdasarkan upah minimum rata-rata dapat dilihat pada Gambar 5.

3) *Mengasumsikan Semua Variabel Selain Variabel  $X_3$*

Variabel selain variabel  $X_3$  diasumsikan konstan, maka berikut adalah persamaan regresi dari angka melek huruf terhadap persentase kasus anak putus sekolah ( $Y$ ).

$$\hat{y} = \begin{cases} -4,1530 + 0,0795x_3 & ; x_3 < 84,43224 \\ 3,3099 - 0,0089x_3 & ; 84,432 \leq x_3 < 97,478 \\ -204,29 + 2,1209x_3 & ; 97,478 \leq x_3 < 97,898 \\ 60,3086 - 0,58196x_3 & ; x_3 \geq 97,89837 \end{cases}$$

Pada interval kurang dari 84,43224 menunjukkan apabila angka melek huruf suatu wilayah kurang dari 84,43224, maka setiap kenaikan angka melek huruf sebesar satu satuan akan meningkatkan persentase anak putus sekolah sebesar 0,0795. Angka melek huruf suatu wilayah yang berada pada interval antara 84,43224 sampai 97,47755 menunjukkan setiap kenaikan angka melek huruf sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,00889. Pada interval antara 97,47755 sampai 97,89837, menunjukkan setiap kenaikan angka melek huruf sebesar satu satuan akan menaikkan persentase anak putus sekolah sebesar 2,1293. Sedangkan pada interval lebih dari 97,89837, menunjukkan setiap kenaikan angka melek huruf sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,58196. Persebaran penduduk berdasarkan angka melek huruf pada Gambar 6.

4) *Mengasumsikan Semua Variabel Selain Variabel  $X_4$*

Dengan mengasumsi variabel selain variabel  $X_4$  konstan, maka berikut adalah persamaan regresi dari rasio guru terhadap murid ( $X_4$ ) terhadap persentase kasus anak putus dari sekolah yang dijalaninya ( $Y$ ).

$$\hat{y} = \begin{cases} -4,1530 - 0,0796x_4 & ; x_4 < 30,89939 \\ -6,9667 + 0,0115x_4 & ; 30,899 \leq x_4 < 51,098 \\ -4,0393 - 0,0458x_4 & ; x_4 \geq 51,09837 \end{cases}$$

Pada interval kurang dari 30,89939 menunjukkan setiap kenaikan rasio guru terhadap murid sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,07958. Wilayah yang termasuk dalam interval pertama adalah Sumenep, Pamekasan, Situbondo, Lamongan dan Kota Madiun. Pada interval antara 30,89939 sampai 51,09837 menunjukkan bahwa setiap kenaikan rasio guru terhadap murid sebesar satu satuan akan menaikkan persentase anak putus sekolah sebesar 0,01148. Sedangkan pada interval lebih dari 51,09837, menunjukkan bahwa setiap kenaikan rasio guru terhadap murid sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,04581. Wilayah yang terdapat dalam interval ketiga ialah Kota Surabaya, Sidoarjo dan Pasuruan. Persebaran penduduk berdasarkan rasio guru terhadap murid pada Gambar 7.

5) *Mengasumsikan Semua Variabel Selain Variabel  $X_5$*

Mengasumsikan semua variabel ( $X_1, X_2, X_3, X_4$ ) selain variabel  $X_5$  konstan, maka persamaan regresi dari rata-rata jumlah anggota keluarga ( $X_5$ ) terhadap persentase kasus anak putus sekolah ( $Y$ ) adalah sebagai berikut.

$$\hat{y} = \begin{cases} -4,15304 - 0,50607x_5 & ; x_5 < 3,86265 \\ -7,74194 - 0,42312x_5 & ; x_5 \geq 3,86265 \end{cases}$$

Pada interval kurang dari 3,86265 menunjukkan bahwa apabila jumlah anggota keluarga di suatu wilayah

kurang dari 3,86265 maka setiap penambahan jumlah anggota keluarga sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,50607. Sedangkan pada interval lebih dari 3,86265, menunjukkan setiap penambahan jumlah anggota keluarga sebesar satu satuan akan menurunkan persentase anak putus sekolah sebesar 0,42312. Wilayah yang terdapat dalam interval kedua ialah Lamongan, Kota Probolinggo, Gresik, Kota Pasuruan, Sampang dan Bangkalan. Persebaran penduduk berdasarkan jumlah anggota keluarga pada Gambar 8.

#### I. Perbandingan Persentase Anak Putus Sekolah Sebelum dan Sesudah Pemodelan

Setelah diperoleh model terbaik yaitu dengan menggunakan kombinasi titik knot 2 3 3 2 1, selanjutnya membandingkan variabel persentase anak putus sekolah di Jawa Timur sebelum dilakukan pemodelan dengan variabel persentase anak putus sekolah di Jawa Timur sesudah pemodelan, perbandingan ini berguna untuk mengetahui bagai mana pengaruh yang diberikan pemodelan dalam persentase anak putus sekolah dengan perbandingan pada Gambar 9.

Gambar 9 menunjukkan perbandingan persentase anak putus sekolah sebelum dan sesudah pemodelan, dimana setelah dilakukan pemodelan persentase anak putus sekolah di beberapa daeran turun seperti persentase anak putus sekolah di Jember Kediri Blitar dan kota/kabupaten lainnya. Maka pemodelan yang dilakukan dengan kombinasi titik knot 2 3 3 2 1 dapat menurunkan persentase anak putus sekolah berdasarkan pola grafik yang diperoleh.

#### V. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan. Tahun 2017 menunjukkan rata-rata banyaknya anak putus sekolah setiap kecamatan/kota di Jawa Timur

sebesar 0,51% artinya setiap 100 siswa usia sekolah 1 orang diantaranya tidak dapat menyelesaikan atau melanjutkan pendidikan, diketahui bahwa Jember memiliki persentase anak putus sekolah tertinggi dan Pacitan dengan persentase anak putus sekolah terendah. Terdapat 22 kabupaten/kota di Jawa Timur yang memiliki persentase anak putus sekolah lebih rendah dari rata-rata dan 16 kabupaten/kota lainnya berada di atas rata-rata. Model terbaik yang dihasilkan untuk memodelkan persentase anak putus sekolah di Jawa Timur tahun 2017 ialah model Regresi Nonparametrik *Spline Truncated* dengan titik kombinasi 2,3,3,2,1 dengan Semua variabel yang digunakan dalam penelitian ini berpengaruh secara signifikan yaitu persentase penduduk miskin, besaran upah minimum rata-rata, angka melek huruf, rasio guru terhadap murid, dan rata-rata jumlah anggota keluarga. Nilai koefisien determinasi dari model tersebut sama dengan 77,67%.

#### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, *Sistem Informasi Rujukan Statistik (SIRUSA)*. Jakarta: Badan Pusat Statistik, 2018.
- [2] Dinas Pendidikan Jawa Timur, *Statistik Pendidikan Jawa Timur 2017*. Surabaya: Pusat Data dan Teknologi Informasi Provinsi Jawa Timur, 2017.
- [3] Organisation for Economic Cooperation and Development (OECD), *Education at Glance 2015*. France: OECD Publishing, 2015.
- [4] UNESCO, *The Drop-Out Problem in Primary Education; Some Case Studies: China, India, Peninsular Malaysia, Socialist Republic of Vietnam, Sri Lanka and Thailand*, 1st ed. Bangkok: UNESCO Regional Office for Education in Asia and the Pacific (Thailand), 1984.
- [5] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd ed. USA: Marcel Dekker, Inc, 1999.
- [6] W. Hardle, *Applied Nonparametric Regression*, 1st ed. New York: The Press Syndicate of The Cambridge University Press, 1990.
- [7] I. N. Budiantara, "Model spline dengan knots optimal," *J. Ilmu Dasar*, vol. 7, no. 2, pp. 77–85, 2006.
- [8] D. N. Gujarati, *Basic Econometrics*, 4th ed. New York: McGraw-Hill, 2004.
- [9] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed., vol. 326. New York: John Wiley & Sons, Inc, 1998.