
PERBANDINGAN REGRESI NONPARAMETRIK KERNEL NWE DAN B-SPLINES PADA PEMODELAN RATA-RATA LAMA SEKOLAH DAN PENGELUARAN PERKAPITA DI INDONESIA

Septie Wulandary¹, Drajat Indra Purnama²

¹ BPS Provinsi Jambi

² BPS Kabupaten Parigi Moutong, Sulawesi Tengah

e-mail: septie@bps.go.id¹, drajatindrapurnama@bps.go.id²

Abstrak

Analisis regresi merupakan salah satu alat statistik yang banyak digunakan untuk mengetahui hubungan antara dua variabel acak atau lebih. Metode penaksiran model regresi terbagi atas regresi parametrik dan nonparametrik. Penelitian ini bertujuan menganalisis pola hubungan pengeluaran perkapita terhadap rata-rata lama sekolah di Indonesia tahun 2018 melalui perbandingan regresi nonparametrik, yaitu regresi kernel dan spline. Regresi kernel yang digunakan adalah regresi kernel dengan metode penaksir Nadaraya-Watson (NWE), sedangkan regresi spline yang digunakan adalah B-Splines. Berdasarkan nilai *Generalized Cross Validation* (GCV) yang minimum dari model regresi B-Splines, digunakan model dengan *degree* 2. Perbandingan model terbaik antara model NWE dan B-Splines dilakukan berdasarkan nilai RMSE terkecil dan kurva yang dihasilkan. Pada penelitian ini, model yang terbaik adalah model B-Splines karena memiliki RMSE 0,705, lebih kecil dibandingkan NWE dengan RMSE 1,854. Selain itu, regresi B-Splines memiliki kurva yang halus dan mengikuti sebaran data dibandingkan kurva NWE.

Kata Kunci: Pengeluaran per Kapita, Rata-rata Lama Sekolah, Regresi Kernel, *B-Splines*

Abstract

Regression analysis is one statistical tool that is widely used to determine the relationship between two or more random variables. The regression model estimation method is divided into parametric and nonparametric regression. This study aims to analyze the pattern of the relationship between per capita expenditure to the mean years school in Indonesia in 2018 through a comparison of nonparametric regression, namely kernel and spline regression. The kernel regression used is kernel regression with the Nadaraya-Watson (NWE) estimator method, while the spline regression used is B-Splines. Based on the minimum Generalized Cross Validation (GCV) value of the B-Splines regression model, a model of degree 2 is used. The best model comparison between the NWE and B-Splines models is based on the smallest RMSE value and the resulting curve. In this study, the best model is the B-Splines model because it has an RMSE of 0.705, smaller than NWE with an RMSE of 1,854. In addition, B-Splines regression has a smooth curve and follows the distribution of data compared to the NWE curve.

Keywords: *Per Capita Expenditure, Mean Years School, Kernel Regression, B-Splines*

1. PENDAHULUAN

Kemiskinan merupakan salah satu indikator keberhasilan pembangunan. Informasi kemiskinan dibutuhkan pemerintah dalam rangka perencanaan kebijakan untuk pengentasan masalah kemiskinan. Informasi mengenai kemiskinan di Indonesia diperoleh dari Badan Pusat Statistik (BPS). BPS menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*) dalam mengukur kemiskinan. Kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran (BPS 2018). Semakin tinggi daya beli masyarakat menunjukkan meningkatnya kemampuan dalam memenuhi kebutuhan hidupnya dan selanjutnya akan berdampak meningkatnya kesejahteraan masyarakat. Indikator kemiskinan (persentase penduduk miskin) diukur berdasarkan rata-rata pengeluaran perkapita. Penduduk yang memiliki rata-rata pengeluaran perkapita per bulan di bawah garis kemiskinan dikategorikan sebagai penduduk miskin (BPS 2018).

Rata-rata pengeluaran per kapita merupakan data yang sangat strategis. Data ini merupakan salah satu komponen penting dalam penghitungan angka kemiskinan suatu wilayah. Data pengeluaran per kapita provinsi atau kabupaten dapat digunakan untuk pengelompokan provinsi atau kabupaten di suatu wilayah menurut kelompok pengeluaran. Hal ini dapat dimanfaatkan oleh pemerintah daerah untuk menyusun kebijakan dan pembangunan terhadap wilayah yang termasuk ke dalam kelompok pengeluaran per kapita yang rendah.

Selain rata-rata pengeluaran per kapita, pendidikan berkualitas merupakan indikator kemiskinan. Indikator dari pendidikan berkualitas antara lain rata-rata lama sekolah, angka partisipasi kasar (APK), dan angka melek huruf. Literatur lain menjelaskan bahwa pendidikan mempunyai peran yang penting dalam pemenuhan kebutuhan dasar lain, peningkatan kualitas hidup, peningkatan pendapatan serta penurunan tingkat kemiskinan (Hudoyo 2017). Rata-rata lama sekolah menunjukkan jenjang pendidikan yang pernah atau sedang diduduki, sehingga semakin tinggi rata-rata lama sekolah maka semakin lama atau tinggi jenjang pendidikan yang ditamatkan. Rata-rata lama sekolah bertujuan untuk melihat kualitas penduduk dalam hal mengenyam pendidikan formal (BPS 2011).

Keterkaitan antara rata-rata lama sekolah dengan tingkat rata-rata pengeluaran perkapita dapat ditunjukkan berdasarkan suatu model regresi. Metode penaksiran model regresi yang diasumsikan mengikuti bentuk persamaan regresi tertentu seperti linier, kuadratik dan yang lainnya disebut regresi parametrik. Akan tetapi seringkali ditemui kesulitan dalam menentukan model melalui sebaran data, sehingga metode yang digunakan adalah regresi nonparametrik. Beberapa teknik penaksiran dalam regresi nonparametrik adalah regresi kernel dan regresi spline yang merupakan teknik penaksiran pada *fitting* kurva regresi yang memperhitungkan penghalusan.

2. METODE PENELITIAN

2.1 Data Penelitian

Penelitian ini menggunakan data yang bersumber publikasi Badan Pusat Statistik (BPS) tahun 2018. Variabel dalam penelitian ini adalah Rata-Rata Lama Sekolah (RLS) sebagai variabel respon dan Pengeluaran Per Kapita (PKP) sebagai variabel prediktor. Pengeluaran Per Kapita (PKP) adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan baik yang berasal dari pembelian, pemberian maupun produksi sendiri dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut.

Sedangkan Rata-rata Lama Sekolah (RLS) didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal.

2.2 Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Menentukan variabel respon dan variabel prediktor.
2. Membuat *scatter plot*.
3. Menentukan prespesifikasi model.
4. Menentukan knot optimum dan parameter penghalus.
5. Membuat *fitting curve*.
6. Menentukan model regresi nonparametrik terbaik

Pengolahan data dan analisis pada penelitian ini menggunakan *Software R*.

2.3 Regresi Nonparametrik

Analisis regresi merupakan salah satu alat statistik yang digunakan untuk mengetahui hubungan antara dua variabel acak atau lebih. Misalkan Y adalah variabel respon dan X adalah variabel prediktor dengan $\{(x_i, y_i), i = 1, 2, \dots, n\}, x_i \in X, y_i \in Y$. Hubungan antara x_i dan y_i dapat diasumsikan mengikuti model regresi sebagai berikut:

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

dengan ε_i adalah *error* acak dan $f(x_i)$ adalah fungsi regresi. Jika $f(x_i)$ tidak diketahui bentuk kurvanya maka pendekatan yang digunakan adalah regresi nonparametrik. Regresi nonparametrik merupakan pendekatan yang sangat fleksibel (Härdle 1994). Hal ini karena regresi nonparametrik membiarkan data sendiri yang akan mencari bentuk estimasi dari fungsi $f(x_i)$ tanpa harus dipengaruhi oleh faktor subyektifitas peneliti (Eubank 1999).

2.4 Penaksir Kernel

Fungsi penaksir Kernel (K) digunakan sebagai salah satu pendekatan mengestimasi kurva regresi nonparametrik. Metode ini sering digunakan karena memiliki bentuk yang lebih fleksibel dan penghitungan matematisnya mudah dikerjakan (Istiqomatul Fajriyah Yuliati and Sihombing 2020). Penaksir densitas kernel memberikan sebaran peluang untuk setiap observasi yang tidak selalu merata dalam interval yang tetap, tetapi secara halus di sekitar observasi. Bentuk bobot kernel ditentukan oleh fungsi Kernel (K), sedangkan ukuran bobotnya ditentukan oleh parameter pemulus h yang disebut *bandwidth*.

Secara umum Kernel (K) dengan *bandwidth* (h) didefinisikan sebagai berikut:

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right) \quad (1)$$

untuk $-\infty < u < \infty, h > 0$

Serta memenuhi:

- i. $K(u) \geq 0$, untuk semua u
- ii. $\int_{-\infty}^{\infty} K(u) du = 1$
- iii. $\int_{-\infty}^{\infty} u^2 K(u) du = \sigma^2 > 0$
- iv. $\int_{-\infty}^{\infty} u K(u) du = 0$

Maka estimator densitas kernel untuk fungsi densitas $f(x)$ adalah:

$$\hat{f}_h = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2)$$

Berdasarkan persamaan (2) terlihat bahwa $f_h(x)$ tergantung pada fungsi kernel K dan parameter h .

2.5 Regresi Kernel

Regresi kernel adalah teknik statistik nonparametrik untuk menaksir nilai ekspektasi bersyarat suatu variabel random. Tujuan regresi kernel adalah mendapatkan hubungan nonlinier antara X dengan Y . Penerapan regresi kernel akan lebih ditekankan pada proses penghalusan (*smoothing*). Penghalusan dengan pendekatan kernel yang selanjutnya dikenal sebagai penghalusan kernel (*kernel smoothing*) sangat tergantung pada fungsi kernel dan *bandwidth*. Taksiran kepadatan kernel sangat tergantung pada fungsi kernel yang digunakan. Kriteria pemilihan fungsi kernel yang baik berdasarkan pada resiko kernel minimum yang dapat diperoleh dari kernel optimal atau kernel-kernel dengan variansi minimum. Kernel mempunyai suatu *bandwidth* atau parameter pemulus yang dinyatakan dengan h . Parameter pemulus ini akan mengontrol *degree* pengaruh titik-titik pada penentuan lokasi. Pemilihan *bandwidth* optimum lebih ditekankan pada keseimbangan antara bias dan varians.

Regresi kernel adalah teknik statistika nonparametrik untuk mengestimasi fungsi regresi $m(x)$ pada model regresi nonparametrik $y_i = m(x_i) + \varepsilon_i$. Menurut (Härdle 1994), pada 1964 Nadaraya dan Watson mempublikasikan metode penaksir $m(\cdot)$ yang disebut metode Nadaraya-Watson Estimator (NWE) sebagai berikut:

$$\hat{m}(X) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)} \quad (3)$$

Untuk X fixed, penaksir $\hat{\theta}$ yang meminimumkan:

$$\sum_{i=1}^n (Y_i - \theta)^2 K_h(X_i - x) \quad (4)$$

Memiliki bentuk $\sum_{i=1}^n a_i Y_i$, NWE merupakan minimizer persamaan (4) dengan:

$$a_i = \frac{K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad (5)$$

2.6 Regresi B-Spline

Regresi B-Spline adalah salah satu metode yang digunakan untuk menaksir kurva regresi nonparametrik. Model B-Spline dengan titik knots digunakan untuk menyelesaikan kelemahan model spline pada saat orde yang tinggi, titik knot yang banyak atau knots yang terlalu dekat yang akan membentuk matriks dalam perhitungan yang hampir singular sehingga persamaan normal sulit diselesaikan. Metode yang digunakan dalam menaksir parameter B-Spline adalah Metode Kuadrat Terkecil (*Ordinary Least Square*) yang didefinisikan secara rekursif.

Menurut (Eubank 1999), B-splines merupakan fungsi *piecewise polynomial* dengan *support* lokal untuk *degree* polinomial tertentu. B-splines ke- j dengan *degree* v berdasarkan knot sekuens t_0, \dots, t_u untuk $j = 1, \dots, v + u$. Dinotasikan dengan formula rekursif berikut:

$$B_j(x; v) = \frac{x - t_j}{t_{j+v-1} - t_j} B_j(x; v - 1) + \left(1 - \frac{x - t_{j+1}}{t_{j+v} - t_{j+1}} B_{j+1}(x; v - 1) \right) \quad (6)$$

dengan:

$$B_j(x; v) = \begin{cases} 1 & \text{jika } t_j \leq x \leq t_{j+1} \\ 0 & \text{lainnya.} \end{cases}$$

B-splines dinormalisasi, berarti bahwa:

$$\forall x: \sum_{j=1}^{v+u} B_j(x; v) = 1.$$

Jika terdapat model regresi *mean* berikut:

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n$$

maka selalu mengasumsikan berdistribusi independen dan identik dengan rata-rata nol dan varians σ^2 . Fungsi $f(\cdot)$ diasumsikan sebagai fungsi halus yang dapat diaproksimasi dengan linear kombinasi dari basis B-splines:

$$f(x) \approx \sum_{j=1}^m \alpha_j B_j(x; v)$$

dengan $\alpha = (\alpha_1, \dots, \alpha_m)$ merupakan vektor koefisien dari basis B-splines $B_j(\cdot; v)$ dengan *degree* v dan $u+1$ *equidistant knots* untuk $j = 1, \dots, u + v = m$. Maka $f(x)$ bisa ditaksir menggunakan:

$$\hat{f}(x) = \sum_{j=1}^m \hat{\alpha}_j B_j(x; v) \quad (7)$$

dengan $\hat{\alpha} = \operatorname{argmin}_{\alpha} \left\{ \sum_{i=1}^n (Y_i - \sum_{j=1}^m \alpha_j B_j(x_i; v))^2 \right\}$.

2.7 Pemilihan Titik Knot Optimal

Model regresi spline terbaik tergantung pada titik knot optimal (Eubank 1999). Kriteria yang harus diperhatikan dalam membentuk model regresi B-spline adalah menentukan orde atau *degree* untuk model regresi dan banyaknya knot. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pada data. Jumlah knot yang terlalu banyak cenderung membuat kurva B-spline menjadi *overfit* dan *rigid* (kasar), sedangkan knot yang terlalu kecil membuat kurva menjadi *oversmooth* dan kurang dapat menggambarkan pola sebaran data.

Dalam Eubank (1998), disebutkan bahwa ukuran kinerja penduga fungsi regresi dapat ditentukan berdasarkan *Generalized Cross Validation* (GCV). GCV diperkenalkan oleh (Craven and Wahba 1978) dengan ide dasar dari GCV adalah memodifikasi *Cross Validation* atau CV (Silverman and Green 1993). Kriteria pemilihan parameter. Nilai GCV yang minimum diperoleh dengan rumus yang dituliskan sebagai berikut:

$$GCV(k) = \frac{MSE(GCV)}{(n^{-1} \operatorname{trace}[I - A(k)])^2} \quad (8)$$

dengan:

I : matriks identitas

$A(k)$: matriks $X(X^T V X)^{-1} X^T V$

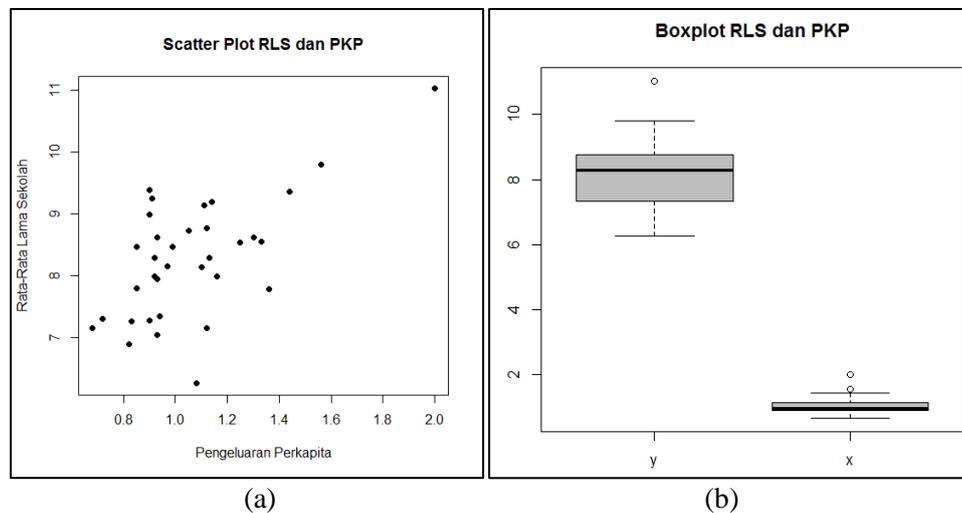
$$MSE(GCV) : \frac{\sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{n} \quad (9)$$

3. HASIL DAN PEMBAHASAN

3.1. Eksplorasi Data

Scatter plot dari data Rata-Rata Lama Sekolah (RLS) sebagai variable respon (Y) dan Pengeluaran Per Kapita (PKP) sebagai variabel terikat (X) di Indonesia tahun 2018 ditunjukkan pada Gambar 1. Berdasarkan *scatter plot* tersebut terlihat bahwa ada kecenderungan apabila pengeluaran perkapita meningkat, maka rata-rata lama sekolah juga meningkat. Hal tersebut dapat juga dijelaskan menggunakan nilai korelasi Pearson sebesar

0,63 yang menunjukkan adanya hubungan yang cukup erat. Akan tetapi, pola hubungan tersebut tidak terlihat linier karena cenderung menyebar. Begitupun apabila dilihat dari *boxplot*, data RLS dan PKP masing-masing memiliki *outlier*.



Gambar 1 (a) *Scatter Plot* Rata-Rata Lama Sekolah (RLS) dan Pengeluaran Per Kapita (PKP) dan (b) *Boxplot* Rata-Rata Lama Sekolah (RLS) dan Pengeluaran Per Kapita (PKP)

Statistik deskriptif dari variabel respon (Y) dan terikat (X) dapat dilihat pada Tabel 1. Pada Tabel 1 terlihat bahwa rata-rata lama sekolah terendah adalah 6,27 tahun yaitu di Provinsi Papua, sedangkan tertinggi yaitu 11,02 tahun yaitu di Provinsi DKI Jakarta. Rata-rata pengeluaran perkapita sebulan terendah adalah di Provinsi Nusa Tenggara Timur, yaitu Rp. 618.483,00, sedangkan tertinggi di Provinsi DKI Jakarta yaitu Rp. 2.000.000,00.

Tabel 1 Statistik Deskriptif RLS (Variabel Y) dan PKP (Variabel X)

Ukuran	Rata-Rata Lama Sekolah (Tahun)	Pengeluaran Perkapita Sebulan (Juta Rupiah)
Min.	6,27	0,68
Q1	7,45	0,90
Median (Q2)	8,29	0,98
Q3	8,75	1,14
Max.	11,02	2,00

3.2. Regresi Kernel

Salah satu kunci utama untuk mendapatkan penghalus yang terbaik pada regresi kernel adalah berdasarkan nilai *bandwidth* h optimum. Pemilihan *bandwidth* (h) merupakan langkah terpenting dalam kernel *smoothing*. Apabila nilai h yang dipilih terlalu kecil akan diperoleh kurva regresi yang sangat kasar (*undersmoothing*), sebaliknya apabila nilai h terlalu besar akan menghasilkan kurva yang sangat mulus (*oversmoothing*). Berdasarkan pengolahan menggunakan *Software R* diperoleh *bandwidth* optimum untuk NWE adalah 0,1761484.

3.3. Regresi B-Spline

Pembentukan model regresi B-Spline memperhatikan tiga kriteria, yaitu menentukan orde untuk model, banyaknya knot, dan lokasi penempatan knot. Knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola pada interval yang berlainan. Untuk memperoleh model B-spline yang optimal (terbaik) maka perlu dipilih lokasi knot yang optimal pula. Kriteria yang digunakan dalam pemilihan knot yang optimal pada penelitian ini yaitu *Generalized Cross Validation* (GCV).

Tabel 2 Nilai GCV dari Knot Optimal Regresi B-Spline

<i>Degree</i>	Jumlah Knot Optimal	Knot Optimal	GCV	RSS
2	1	0,94708	0,63801	16,88837
3	1	1,39311	0,66645	16,48492
4	1	0,88987	0,70343	16,22026

Berdasarkan Tabel 2, pasangan nilai GCV diperoleh jumlah knot optimum sebanyak 1 pada masing-masing *degree* model. Model B-Spline dengan *degree* 2 memiliki nilai GCV minimum merupakan model yang terbaik sehingga akan dilakukan estimasi parameter untuk model dengan *degree* 2 tersebut. Hasil estimasi parameter ditampilkan pada Tabel 3.

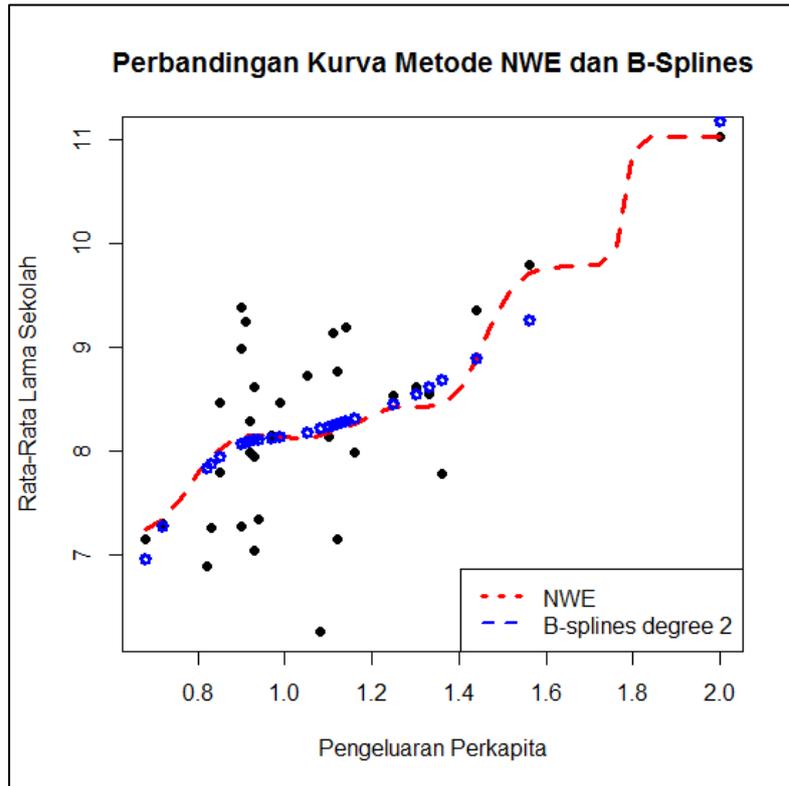
Tabel 3 Hasil Regresi Nonparamterik B-Spline

Koefesien	Estimate	Std. Error	t-Statistik	p-value
(Intercept)	6,9638	0,5987	11,631	1,21e-12 ***
bs(x, knots = c(K), <i>degree</i> = 2)1	1,1043	0,7213	1,531	0,136234
bs(x, knots = c(K), <i>degree</i> = 2)2	1,3618	0,7811	1,743	0,091512 .
bs(x, knots = c(K), <i>degree</i> = 2)3	4,2022	0,9524	4,412	0,000122 ***
F-Statistik : 8,157			p-value : 0,000405	
Multiple R-squared : 0,4492			Adjusted R-squared : 0,3942	

Berdasarkan hasil regresi pada Tabel 3, diperoleh koefisien determinasi sebesar 0,4 artinya variasi rata-rata lama sekolah yang mampu dijelaskan oleh variasi pengeluaran perkapita sebulan dalam model hanya sebesar 40 persen, sisanya dijelaskan oleh variabel lain di luar model. Sedangkan dari model tersebut diperoleh bahwa nilai *p-value* F-statistik=0,000405 < 0,05, hal ini mengindikasikan bahwa model signifikan secara simultan. Sementara hasil uji t-statistik menunjukkan bahwa tidak semua parameter yang signifikan di dalam model.

3.4. Perbandingan Regresi Kernel dan Regresi B-Spline

Perbandingan antara kurva regresi kernel dan kurva regresi B-spline disajikan pada Gambar 2. Pada gambar tersebut menunjukkan bahwa kurva B-Spline lebih *smooth* (halus) dan terlihat lebih mengikuti sebaran data dibandingkan kurva NWE. Sedangkan kurva NWE mengikuti pola data tetapi kurang optimal dalam mencapai titik terendah observasi.



Gambar 2 Perbandingan Regresi NWE dan B-Splines

Untuk melihat bahwa model yang dihasilkan akurat atau tidak dapat dilihat dari nilai RMSE. Semakin kecil nilai RMSE maka dapat dikatakan modelnya semakin baik. Perbandingan nilai RMSE antara model NWE dengan B-spline disajikan pada Tabel 4.

Tabel 4 Perbandingan nilai RMSE Regresi Kernel dan B-Spline

No	Model	RMSE
1	NWE	1,8544
2	B-Spline	0,7048

Berdasarkan hasil pada Tabel 4 diperoleh bahwa nilai RMSE model B-Spline lebih kecil dibandingkan model NWE. Sehingga, pada penelitian ini dapat dinyatakan bahwa pola hubungan antara rata-rata lama sekolah dan pengeluaran perkapita di Indonesia lebih tepat menggunakan model B-Spline *degree 2*.

4. KESIMPULAN

Berdasarkan hasil analisis terhadap data rata-rata lama sekolah dan pengeluaran perkapita dapat disimpulkan bahwa:

1. Terdapat hubungan positif yang cukup kuat antara rata-rata lama sekolah dan pengeluaran perkapita rumah tangga di Indonesia tahun 2018, yaitu apabila pengeluaran per kapita rumah tangga meningkat, maka rata-rata lama sekolah juga akan meningkat.
2. Metode nonparametrik yang paling cocok untuk menggambarkan pola hubungan antara rata-rata lama sekolah dengan pengeluaran perkapita adalah regresi B-Splines *degree 2*, dengan metode optimasi parameter GCV.

3. Perbandingan model terbaik antara model kernel dengan NWE dan B-Spline berdasarkan nilai RMSE terkecil. Model B-Splines memiliki RMSE 0,705, lebih kecil dibandingkan NWE dengan RMSE 1,854, sehingga B-Splines ditetapkan sebagai model terbaik.
4. Regresi B-Splines memiliki kurva yang lebih halus dan mengikuti sebaran data dibandingkan kurva NWE.

DAFTAR PUSTAKA

- BPS (2011), *Ensiklopedia Indikator Ekonomi Sosial*, Jakarta: BPS.
- BPS (2018), *Data dan Informasi Kemiskinan Kabupaten/Kota Tahun 2018*, Jakarta: BPS.
- Craven, P., and Wahba, G. (1978), "Smoothing noisy data with spline functions," *Numerische Mathematik*, 31, 377–403. <https://doi.org/10.1007/BF01404567>.
- Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker.
- Härdle, W. (1994), *Applied Nonparametric Regression*, Berlin: Springer-Verlag.
- Hudoyo, L. . (2017), "Pemodelan Hubungan antara Rata-Rata Lama Sekolah dan Pengeluaran Rumah Tangga Menggunakan Constrained B-Splines (COBS) pada Regresi Kuantil," Universitas Padjajaran.
- Istiqomatul Fajriyah Yuliati, and Sihombing, P. (2020), "Pemodelan Fertilitas Di Indonesia Tahun 2017 Menggunakan Pendekatan Regresi Nonparametrik Kernel dan Spline," *Jurnal Statistika dan Aplikasinya*, 4, 48–60. <https://doi.org/10.21009/JSA.04105>.
- Silverman, B., and Green, P. (1993), *Nonparametrik Regression and Generalized Linear Models (a Roughness Penalty Approach)*, New York: Chapman & Hall.