

PERBANDINGAN REGRESI LINIER BERGANDA DENGAN *SPLINE TRUNCATED*
(STUDI KASUS: KEMISKINAN DI PROVINSI PAPUA)

Comparison of Multiple Linear Regression With Truncated Spline
(Case: Poverty in Papua Province)

Ni Putu Ayu Mirah Mariati^{1*}, I Wayan Sudiarsa², Ni Made Sukma Sanjiwani³

^{1,3} Universitas Mahasaraswati Denpasar,

Jalan Kamboja No.11A, Denpasar, Bali, Indonesia

² Universitas PGRI Mahadewa Indonesia

Jalan Seroja, Tonja, Denpasar Utara, Bali, Indonesia

*Pos-el: ayumirahmariati@unmas.ac.id, wayansudiarsa1804@gmail.com,
sukmasanjiwani@unmas.ac.id

ABSTRACT

Regression analysis was used to study the pattern of the relationship between the response variable and the predictor variable. There are two approaches used in this method, namely parametric and nonparametric. The most popular parametric regression used is multiple linear regression. In addition to parametric regression there are nonparametric regression approaches. One of the well-known nonparametric regression is Spline Truncated. This research to compare the multiple linear regression method with the Spline Truncated in the case of poverty in Papua Province. Based on the results of research on poverty cases in Papua Province, it can be concluded that the Spline Truncated regression is better than the multiple linear regression with R^2 of 88.39%.

Keywords: Regression, Parametric, Nonparametric, Poverty.

ABSTRAK

Analisis regresi digunakan untuk mempelajari tentang pola hubungan antara variabel respon dengan variabel prediktor. Terdapat dua pendekatan yang digunakan dalam metode tersebut yaitu parametrik dan nonparametrik. Regresi parametrik yang populer digunakan adalah Regresi Linier Berganda. Selain regresi parametrik ada pendekatan regresi nonparametrik. Salah satu regresi nonparametrik yang terkenal yaitu *Spline Truncated*. Penelitian ini bertujuan untuk membandingkan antara metode Regresi Linier Berganda dengan *Spline Truncated* pada kasus kemiskinan di Provinsi Papua. Berdasarkan hasil penelitian pada kasus kemiskinan di Provinsi Papua dapat disimpulkan bahwa regresi *Spline Truncated* lebih baik dibandingkan dengan Regresi Linier Berganda dengan R^2 sebesar 88,39%.

Kata Kunci: Regresi, Parametrik, Nonparametrik, Kemiskinan.

PENDAHULUAN

Analisis regresi merupakan metode analisis data yang menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor. Menurut

Hardle pada tahun 1990 dalam mengestimasi kurva regresi ada dua pendekatan yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik. Pendekatan kurva regresi yang sering digunakan adalah pendekatan

regresi parametrik, dimana diasumsikan bentuk kurva regresi diketahui (Draper dan Smith, 1992). Salah satu pendekatan parametrik yang populer adalah regresi linier. Berikut ini penelitian mengenai regresi linier yaitu meneliti selang kepercayaan dalam regresi linier sederhana dilakukan oleh Liu dkk., pada tahun 2007. Selanjutnya, Grömping pada tahun 2009 melakukan penelitian simulasi terhadap regresi linier sederhana yang membandingkan dengan regresi pohon. Penelitian estimator regresi linier sederhana menggunakan *measure of location-based estimators* dilakukan oleh Liu dkk., pada tahun 2016.

Selain pendekatan parametrik, terdapat pendekatan nonparametrik. Pendekatan nonparametrik yang sering digunakan oleh para peneliti salah satunya yaitu Spline. Spline merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Salah satu kelebihan Spline adalah model yang cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak (I. K. Sukendra, 2020). Kelebihan ini terjadi karena dalam Spline terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Sukendra et al., 2022). Dengan titik knot ini, Spline dapat memberikan fleksibilitas

yang lebih baik daripada polinomial, sehingga memungkinkan untuk menyesuaikan diri secara efektif terhadap karakteristik lokal dari suatu fungsi atau data. Salah satu basis fungsi yang biasa digunakan dalam pendekatan spline, yaitu *Spline Truncated* (Dani dkk., 2020). Adapun beberapa penelitian sebelumnya yang menggunakan pendekatan regresi nonparametrik *Spline Truncated* diantaranya (Saputro dkk., 2018)

Pendekatan regresi parametrik dan nonparametrik dapat diterapkan di berbagai bidang keilmuan, salah satunya yang berkaitan dengan masalah sosial. Salah satu masalah yang sering dijumpai yaitu kemiskinan. Berdasarkan data BPS pada Tahun 2020, Provinsi Papua memiliki persentase penduduk miskin tertinggi di Indonesia. Badan Pusat Statistik (BPS) mendefinisikan kemiskinan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak. Berdasarkan uraian diatas, maka pada penelitian ini akan dimodelkan persentase kemiskinan Kabupaten/Kota di Provinsi Papua dengan menggunakan Regresi Linier Berganda dengan *Spline Truncated*. Diharapkan hasil penelitian ini dapat bermanfaat dalam pengembangan ilmu statistika khususnya pada bidang regresi.

Tujuan penelitian yaitu membandingkan antara Regresi Linier Berganda dengan regresi *Spline Truncated* pada kasus kemiskinan di Provinsi Papua. Selanjutnya menginterpretasikan model terbaik.

METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder tahun 2020 yang diperoleh dari publikasi Badan Pusat Statistik (BPS) Provinsi Papua Unit observasi yang digunakan dalam penelitian ini adalah 29 Kabupaten/Kota yang ada di Provinsi Papua.

Variabel yang digunakan dalam penelitian ini adalah y adalah persentase penduduk miskin, x_1 adalah angka harapan hidup, x_2 adalah tingkat pengangguran terbuka, x_3 adalah angka melek huruf, x_4 adalah laju pertumbuhan PDRB (Widana et al., 2020).

Langkah-langkah aplikasi pada data kemiskinan di Provinsi Papua:

- Membuat *Scatterplot*.
- Memodelkan menggunakan Regresi Linier Berganda.
- Menghitung nilai R^2 model Regresi Linier Berganda
- Memodelkan data dengan menggunakan *Spline Truncated*.

- Memilih titik knot optimal dengan metode GCV.
- Menghitung nilai R^2 model *Spline Truncated* yang terbaik.
- Membandingkan model Regresi Linier Berganda dengan *Spline Truncated*.

Membuat interpretasi model.

HASIL DAN PEMBAHASAN

Penerapan dari model Regresi Linier Berganda dan *Spline Truncated* pada data kemiskinan di Provinsi Papua. Penerapan dilakukan berdasarkan tahapan penelitian yang diberikan dalam metode penelitian. Berikut ini dijelaskan penerapan menggunakan Regresi Linier Berganda.

Regresi Linier Berganda

Dari hasil regresi, maka didapatkan koefisien regresi yang dapat dilihat pada

Tabel 1.

Tabel 1
Hasil Uji Regresi (Koefisien Regresi)

Variabel	Koefisien	P_{value}
x_1	-0,06	0,04
x_2	-1,12	0,04
x_3	-0,22	0,03
x_4	-0,35	0,27

Berdasarkan tabel diatas maka didapatkan persamaan Regresi Linier Berganda sebagai berikut:

$$\hat{y} = 53,91 - 0,06x_1 - 1,12x_2 - 0,22x_3 - 0,35x_4$$

Pada uji serentak model Regresi Linier Berganda di dapat $P_{value} = 0,00$, dengan $\alpha = 5\%$, maka. Sehingga secara serentak variabel x_1 , x_2 , x_3 dan x_4 berpengaruh terhadap y . Uji individu terlihat pada **Tabel 1** bahwa keempat variabel berpengaruh signifikan terhadap presentase kemiskinan.

Hasil Uji Koefisien Determinasi di dapat $R^2 = 53,6\%$. Dapat dikatakan bahwa 53,6% presentase kemiskinan dapat dijelaskan oleh keempat variabel prediktor.

Regresi *Spline Truncated*

Pada pembahasan ini akan dilakukan pemodelan mengenai variabel-variabel yang diduga mempengaruhi kemiskinan. Metode yang digunakan untuk memodelkan data ini adalah regresi nonparametrik *Spline Truncated*. Berikut adalah model matematis dari regresi nonparametrik *Spline Truncated*.

$$y_i = \sum_{j=1}^p \left(\sum_{k=1}^q \hat{\beta}_{kj} x_{ki}^v + \sum_{l=1}^r \hat{\beta}_{j(l+q)} (x_{ji} - K_{jl})_+^q \right)$$

Titik knot merupakan titik dimana pola dari data berubah. Untuk mendapatkan titik knot optimal maka menggunakan metode GCV. Memilih nilai knot yang paling optimal digunakan nilai GCV yang paling minimum. Titik knot yang digunakan dalam penelitian ini

adalah satu knot, dua knot, tiga knot dan kombinasi knot. Berikut ini akan ditampilkan perolehan knot optimal.

Tabel 2
Perbandingan Nilai GCV

Knot	GCV
1 Knot	43,50
2 Knot	41,09
3 Knot	38,43
Knot Kombinasi	32,47

Berdasarkan **Tabel 2** diperoleh knot optimal yaitu knot kombinasi. Kombinasi knot adalah kombinasi antara satu titik knot, dua titik knot, dan tiga titik knot. Kombinasi ini digunakan untuk memilih titik knot optimal. Dalam pemilihan titik knot optimal pada model regresi nonparametrik *Spline Truncated* dengan **kombinasi knot** ini dipilih nilai GCV yang paling minimum.

Dalam mendapatkan model regresi nonparametrik *Spline Truncated* yang terbaik adalah dengan menggunakan titik knot optimal. Dari hasil pemilihan titik knot optimal yang dilakukan, maka model regresi dengan menggunakan kombinasi titik knot adalah yang terbaik. Hasil dari estimasi parameter dengan menggunakan kombinasi titik knot adalah sebagai berikut.

$$\hat{y} = 98,83 - 0,72 x_1 - 48,38 (x_1 - 71,66)_+ - 8,28 x_2 + 41,92 (x_2 - 2,84)_+ - 68,14 (x_2 - 3,72)_+ + 31,57 (x_2 - 4,59)_+ +$$

$$\begin{aligned}
& -0,02 x_3 - 3,56 (x_3 - 96,26)_+ + \\
& -1,12 x_4 + 2,63 (x_4 - 0,04)_+ + \\
& -12,47 (x_4 - 1,60)_+ + 14,28 (x_4 - 3,16)_+.
\end{aligned}$$

Model regresi *Spline Truncated* dengan kombinasi titik knot ini memiliki R^2 sebesar 88,39%. Hal ini memiliki arti bahwa model ini dapat menjelaskan kemiskinan sebesar 88,39%.

Intepretasi Model Terbaik

Dari hasil analisis, model regresi nonparametrik *Spline Truncated* terbaik adalah dengan kombinasi titik knot. Model terbaik dapat diinterpretasikan sebagai berikut.

1. Apabila variabel x_2 , x_3 dan x_4 dianggap konstan, maka besar pengaruh persentase angka harapan hidup (x_1) terhadap presentase kemiskinan (y) adalah

$$\begin{aligned}
\hat{y} &= -0,72 x_1 - 48,38 (x_1 - 71,66)_+ + \\
&= \begin{cases} -0,72 x_1; & x_1 < 71,66 \\ -49,1 x_1 + 3.466,9; & x_1 \geq 71,66 \end{cases}
\end{aligned}$$

Dari model tersebut dapat diinterpretasikan yaitu Apabila presentase angka harapan hidup lebih dari 71,66 dan apabila presentase angka harapan hidup naik 1%, maka presentase kemiskinan akan turun sebesar 49,1. Wilayah yang termasuk dalam interval ini adalah Kabupaten Mimika.

2. Apabila variabel x_1 , x_3 dan x_4 dianggap konstan, maka besar pengaruh tingkat pengangguran terbuka (x_2) terhadap presentase kemiskinan (y) adalah

$$\begin{aligned}
\hat{y} &= -8,28 x_2 + 41,92 (x_2 - 2,84)_+ + \\
& - 68,14 (x_2 - 3,72)_+ + 31,57 (x_2 - 4,59)_+ + \\
&= \begin{cases} -8,28 x_2; & x_2 < 2,84 \\ 33,64 x_2 - 119,05; & 2,84 \leq x_2 < 3,72 \\ -34,5 x_2 + 134,43; & 3,72 \leq x_2 < 4,59 \\ -2,93 x_2 - 10,48; & x_2 \geq 4,59 \end{cases}
\end{aligned}$$

Dari model tersebut dapat diinterpretasikan yaitu apabila tingkat pengangguran terbuka kurang dari 2,84 dan apabila tingkat pengangguran turun 1%, maka presentase kemiskinan akan cenderung turun sebesar 8,28. Wilayah yang termasuk dalam interval ini adalah Jayawijaya, Paniai, Puncak Jaya, Asmat, Tolikara, Keerom, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya dan Deiyai. Selanjutnya apabila tingkat pengangguran terbuka antara 3,72 hingga 4,59 dan apabila tingkat pengangguran terbuka turun 1%, maka presentase kemiskinan turun sebesar 34,5. Wilayah yang masuk dalam interval ini adalah Kabupaten Yahukimo, Pegunungan Bintang dan Supiori. Apabila tingkat pengangguran terbuka lebih dari 4,59 dan apabila tingkat pengangguran

terbuka turun 1%, maka presentase kemiskinan akan berkurang sebesar 2,93. Wilayah yang termasuk dalam interval ini adalah Kota Jayapura, Jayapura, Biak, Mimika, Boven Digoel, Mappi, Sarmi, Keerom dan Waropen.

3. Apabila variabel x_1 , x_2 dan x_4 , dianggap konstan, maka besar pengaruh persentase angka melek huruf (x_3) terhadap presentase kemiskinan (y) adalah

$$\hat{y} = -0,02 x_3 - 3,56 (x_3 - 96,26)_+ + \begin{cases} -0,02 x_3; & x_3 < 96,26 \\ -3,58 x_3 + 342,69; & x_3 \geq 96,26 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu apabila angka melek huruf lebih dari 96,26, apabila persentase angka melek huruf naik 1%, maka presentase kemiskinan akan cenderung turun sebesar 3,58. Wilayah yang termasuk dalam interval ini adalah kabupaten Jayapura, Nabire, Biak Numfor, Mimika, Sarmi, Waropen, Supiori, Kota Jayapura. Wilayah lain di Provinsi Papua memiliki angka melek huruf dibawah 96,26. Apabila angka melek huruf naik 1% maka akan menurunkan presentase kemiskinan sebesar 0,02.

4. Apabila variabel x_1 , x_2 dan $s.x_3$ dianggap konstan, maka besar laju

pertumbuhan PDRB (x_4) terhadap presentase kemiskinan (y) adalah

$$\hat{y} = -1,12x_4 + 2,63(x_4 - 0,04)_+ - 12,47(x_4 - 1,60)_+ + 14,28(x_4 - 3,16)_+ = \begin{cases} -1,12x_4; & x_4 < 0,04 \\ 1,51x_4 - 0,11; & 0,04 \leq x_4 < 1,60 \\ -10,96x_4 + 20,01; & 1,60 \leq x_4 < 3,16 \\ 3,32x_4 - 25,11; & x_4 \geq 3,16 \end{cases}$$

Dari model tersebut dapat diinterpretasikan yaitu Apabila laju pertumbuhan PDRB kurang dari 0,04 dan laju pertumbuhan PDRB naik 1 satuan, maka presentase kemiskinan akan turun sebesar 1,12. Wilayah yang termasuk dalam interval ini adalah Kabupaten Jayawijaya, Jayapura, Kepulauan Yapen, Biak Numfor, Puncak Jaya, Boven Digoel, Waropen dan Kota Jayapura. Apabila laju pertumbuhan PDRB antara 0,04 hingga 1,60, apabila laju pertumbuhan PDRB naik 1%, maka presentase kemiskinan akan cenderung turun sebesar 1,51. Wilayah yang termasuk dalam interval ini adalah Kabupaten Merauke, Nabire dan Mappi.

SIMPULAN DAN SARAN

Berdasarkan pemodelan yang telah dilakukan menggunakan Regresi Linier Berganda dan *Spline Truncated* pada kasus kemiskinan di Provinsi Papua maka dapat disimpulkan bahwa model *Spline Truncated* lebih baik daripada Regresi Linier Berganda. Nilai R^2 pada *Spline Truncated*

yaitu sebesar 88,39% sedangkan R^2 nilai pada Regresi Linier Berganda sebesar 53,6%.

Variabel-variabel yaitu angka harapan hidup, tingkat pengangguran terbuka, angka melek huruf, laju pertumbuhan PDRB memiliki pengaruh yang signifikan terhadap presentase kemiskinan di Provinsi Papua.

DAFTAR PUSTAKA

- Ali, M. (2013). The Semitization of Itihasa: Intertextuality of the Mahabharata and the Ramayana in the Judeo-Islamic texts. *Atavisme: Jurnal Ilmiah Kajian Sastra*, 16(1), 1-13.
- Draper, N.R., dan Smith, H. (1998). *Applied Regression Analysis, Third Edition*. John Wiley & Sons, USA.
- Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician*, 63(4), (308–319), doi:10.1198/tast.2009.08199.
- Dani, A. T. R. Adrianingsih, N. Y. dan Ainurrochmah, A. (2020) Pengujian Hipotesis Simultan Model Regresi Nonparametrik *Spline Truncated* dalam Pemodelan Kasus Ekonomi: *Journal of Probability and Statistics*, 01, (98-106).
- Hardle, W., (1990), *Applied Nonparametric Regression*, New York: Cambridge University Press.
- Liu, D., Lin, X., dan Ghosh, D. (2007). Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models: *Biometrics*, 63, (1079–1088), doi:10.1111/j.1541-0420.2007.00799.x.
- Liu, X., and Preve, D., (2016). Measure of Location-based Estimators in Simple Linear Regression: *Journal of Statistical Computation and Simulation*, 86(9), (1771–1784). doi:10.1080/00949655.2015.1082131.
- Saputro, D. R. S., Demu, K. R. dan Widyaningsih, P. (2018). Nonparametric Truncated Spline Regression Model on Data of Human Development Index (HDI) in Indonesia: *IOP Conf. Series: Journal of Physics*, 1188.
- Sukendra, I. K. (2020). Developing teaching materials for Trigonometry in mathematics with realistic orientation using HOTS questions. *Journal of Physics: Conference Series*, 1663(1). <https://doi.org/10.1088/1742-6596/1663/1/012020>
- Sukendra, I. K., Suharta, I. G. P., Ardana, I. M., & Ariawan, P. W. (2022). *The Mechanism Development of Digital Mathematics Material Study Based on STEM*. 7(2), 4098–4104. https://kalaharijournals.com/resources/FebV7_I2_495.pdf
- Widana, I. W., Sumandya, I. W., Sukendra, K., & Sudiarsa, I. W. (2020). Analysis of Conceptual Understanding, Digital Literacy, Motivation, Divergent of Thinking, and Creativity on the Teachers Skills in Preparing Hots-based Assessments. *Journal of Advanced Research in Dynamical and Control Systems*, 12(8), 459–466. <https://doi.org/10.5373/jardcs/v12i8/20202612>