



Regresi Nonparametrik Kernel Gaussian pada Pemodelan Angka Kelahiran Kasar di Provinsi Nusa Tenggara Barat

Deni Pratiwi^{a,*}, Lalu Abd Azis Mursy^b, Muhammad Rizaldi^c, Nurul Fitriyani^d

^{a,*}Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No 62, Mataram, Indonesia, 83125. Email: denipratiwi26@gmail.com

^bProgram Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No 62, Mataram, Indonesia, 83125. Email: azizazure4@gmail.com

^cProgram Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No 62, Mataram, Indonesia, 83125. Email: mrizaldi787@gmail.com

^dProgram Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jalan Majapahit No 62, Mataram, Indonesia, 83125. Email: nurul.fitriyani@unram.ac.id

ABSTRACT

This study aims to model Crude Birth Rates (CBR) in West Nusa Tenggara Province. The nonparametric regression method was used in this research by considering data distribution patterns that do not show a linear relationship between variables. In this case, the kernel nonparametric regression using the Gaussian function and the Nadaraya-Watson estimator. The results showed optimal bandwidths of 0.55542837, 1.29042927, 0.94706041, and 0.92278896 with a value of minimum Generalized Cross-Validation (GCV) of 0.000000000432613511, which was minimized by the simulated annealing algorithm. The resulting model's accuracy can be seen from the coefficient of determination (R^2) of 99.23% and the Mean Absolute Percentage Error (MAPE) of 0.007049%.

Keywords: bandwidth, Generalized Cross-Validation, MAPE, Nadaraya-Watson estimator, simulated annealing algorithm

ABSTRAK

Penelitian ini bertujuan untuk memodelkan Angka Kelahiran Kasar (AKK) di Provinsi Nusa Tenggara Barat. Penelitian ini menggunakan metode regresi nonparametrik dengan mempertimbangkan pola distribusi data yang tidak menunjukkan hubungan linear antar variabel. Dalam hal ini digunakan regresi nonparametrik kernel dengan fungsi Gaussian dan estimator Nadaraya-Watson. Hasil penelitian menunjukkan *bandwidth* optimal sebesar 0,55542837, 1,29042927, 0,94706041, dan 0,92278896, dengan *Generalized Cross-Validation* (GCV) minimum 0.000000000432613511 yang diminimumkan dengan algoritma *simulated annealing*. Ketepatan model yang

* Corresponding author.

Alamat e-mail: denipratiwi26@gmail.com

dihasilkan dapat dilihat dari nilai koefisien determinasi sebesar 99,23% dan *Mean Absolut Percentage Error* (MAPE) sebesar 0.007049%.

Kata Kunci: *bandwidth*, *Generalized Cross-Validation*, MAPE, estimator Nadaraya-Watson, algoritma *simulated annealing*

Diserahkan: 11-10-2020; Diterima: 25-12-2020;

Doi: <https://10.29303/emj.v3i2.78>

1. Pendahuluan

Indonesia merupakan negara dengan jumlah penduduk terbesar keempat di dunia. Hal ini disebabkan karena pertambahan jumlah penduduk yang masih tergolong tinggi. Pertambahan jumlah penduduk tidak dapat terlepas dari faktor demografi, diantaranya adalah fertilitas (kelahiran), mortalitas (kematian), dan migrasi (perpindahan penduduk). Suatu ciri bahwa suatu negara sedang berkembang yaitu memiliki tingkat fertilitas atau kelahiran yang lebih tinggi jika dibandingkan dengan negara maju. Chodijah (2011) menyatakan salah satu indikator sosial demografi yang penting dalam mempengaruhi fertilitas adalah usia perkawinan pertama, karena usia perkawinan pertama wanita berkaitan dengan permulaan wanita melakukan hubungan suami istri yang memungkinkan wanita berisiko untuk hamil. Oleh karena itu, pada masyarakat yang wanitanya melakukan perkawinan pertama pada usia muda, angka kelahirannya juga lebih tinggi dibandingkan dengan masyarakat yang wanitanya melakukan perkawinan pertama kali pada usia lebih dewasa.

Data sensus tahun 2010 menunjukkan bahwa masih terdapat beberapa provinsi yang usia perkawinan pertamanya di bawah 20 tahun, salah satunya yaitu provinsi Nusa Tenggara Barat. Berdasarkan data Pemerintah Provinsi dalam NTB Satu Data, kasus pernikahan dini anak usia 10-19 tahun di 10 kabupaten/kota di NTB berada di atas 30 persen sampai 58 persen pada 2018. Seperti Lombok Barat dengan 49,89 persen, Lombok Tengah 57,98 persen, Lombok Timur 58,05 persen. Kemudian Lombok Utara 47,95 persen, Sumbawa Barat 37,81 Persen, Sumbawa 37,32 persen, Dompu 34,19 persen, Bima 34,64 persen, Kota Mataram 42,14 persen dan Kota Bima 32,95 persen. Secara garis besar, umur perkawinan pertama memiliki pengaruh yang cukup kuat terhadap tingkat fertilitas seorang wanita (Jones dan Gubhaju, 2008).

Selain dari faktor sosial demografi di atas, fertilitas juga dipengaruhi oleh faktor non-demografi, diantaranya adalah tingkat pendidikan, keadaan

ekonomi penduduk, urbanisasi, industrialisasi, sosial dan budaya. Fadlyana (2009) menyatakan faktor pendidikan merupakan faktor yang berpengaruh terhadap usia perkawinan pertama, semakin muda usia menikah, maka semakin rendah tingkat pendidikan yang dicapai. Faktor pendidikan juga sangat berpengaruh terhadap angka fertilitas, semakin tinggi pendidikan, maka jumlah usia perkawinan muda rendah. Akibatnya tingkat fertilitas menjadi berkurang.

Salah satu cara untuk mengetahui hubungan fungsional antara angka kelahiran kasar dengan faktor-faktor yang mempengaruhinya adalah dengan menggunakan analisis regresi. Analisis awal menunjukkan bahwa tidak terdapat bentuk pola khusus pada plot data, sehingga pendekatan yang digunakan adalah regresi dengan pendekatan nonparametrik. Salah satu pendekatan dalam regresi nonparametrik yang sering digunakan adalah pendekatan kernel (Hardle, 1994). Penggunaan metode regresi nonparametrik kernel berdasarkan pada analisis awal mengenai karakteristik data yang tidak menunjukkan pola hubungan yang mudah untuk digambarkan atau tidak diketahui.

Terdapat beberapa jenis fungsi kernel, seperti kernel seragam, kernel segitiga, kernel kuartik, kernel Gaussian, dan lain-lain. Salah satu fungsi yang sering digunakan adalah fungsi kernel Gaussian. Fungsi ini identik dengan fungsi kepadatan peluang normal standar (Hardle, et al., 2004).

Menurut Hardle (1994), pemilihan fungsi kernel sebenarnya tidak terlalu berpengaruh terhadap nilai prediksinya, dan yang paling berpengaruh adalah pemilihan *bandwidth*. *Bandwidth* dalam regresi nonparametrik kernel merupakan suatu parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *undersmoothing* yaitu sangat kasar, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *oversmoothing* yaitu sangat mulus tetapi tidak sesuai dengan pola data. Oleh karena itu, perlu dipilih

bandwidth optimal untuk mendapatkan kurva yang optimal.

Oleh karena itu, berdasarkan data dan fakta yang telah dijelaskan, peneliti tertarik untuk mengetahui model angka kelahiran kasar di provinsi Nusa Tenggara Barat dengan menggunakan regresi nonparametrik kernel.

2. Metode

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistika (BPS) Provinsi Nusa Tenggara Barat, yaitu data hasil SUSENAS 2018 dari 10 Kabupaten/Kota di Provinsi Nusa Tenggara Barat, serta dengan data tambahan dari Dinas Kesehatan Provinsi Nusa Tenggara Barat. Penelitian ini menggunakan beberapa variabel, seperti yang tertera pada Tabel 1.

Tabel 1 – Variabel-variabel penelitian

Variabel	Jenis Variabel
Angka Kelahiran Kasar (AKK)	Respon (Y)
Rata-rata usia kawin pertama perempuan 10 tahun ke atas	Prediktor (X_1)
Persentase penduduk usia 10 tahun ke atas yg sudah menikah	Prediktor (X_2)
Persentase perempuan usia 15 tahun yang tidak/belum tamat SD	Prediktor (X_3)
Persentase perempuan usia 15 tahun keatas yang tamat universitas/sarjana	Prediktor (X_4)

Penelitian dilakukan dengan tahapan sebagai berikut: (1) studi literatur; (2) mengumpulkan data; (3) menentukan statistika deskriptif data; (4) menghitung korelasi Pearson dan membuat scatter plot antara variabel respon dengan masing-masing variabel prediktor; (5) menentukan *bandwith* optimal dengan menggunakan *simulated annealing algorithm* serta memperhatikan nilai GCV; (6) menentukan model regresi nonparametrik kernel dengan fungsi Gaussian; (7) melakukan uji asumsi residual regresi nonparametrik; dan (8) membuat kesimpulan (Budiantara, 2015).

3. Hasil dan Pembahasan

Hasil dan Penelitian ini diawali dengan menentukan statistik deskriptif data seperti yang tertera pada Tabel 2.

Tabel 2 – Statistika deskriptif data

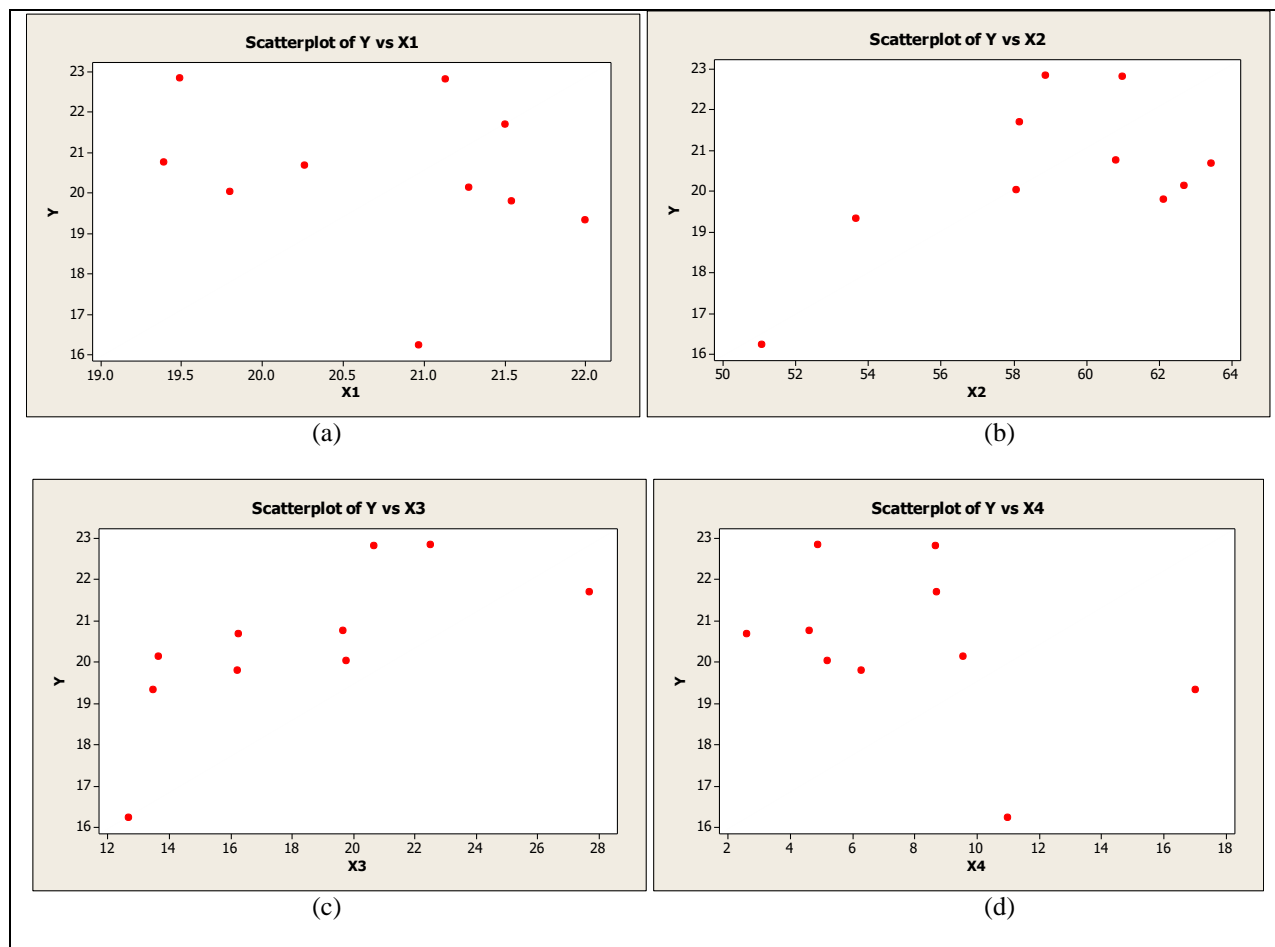
Variabel	Min	Maks	Rata-rata	Standar Deviasi
Y	16,23	22,83	20,4280	1,90588
X_1	22,83	22,00	20,7360	0,93099
X_2	51,08	63,45	58,9960	3,99339
X_3	12,69	27,70	18,2640	4,71594
X_4	2,59	17,01	7,8470	4,1446

Setelah menentukan statistik deskriptif data selanjutnya menentukan korelasi Pearson variabel respon terhadap masing-masing variabel prediktor. Nilai korelasi yang dihasilkan dapat dilihat pada Tabel 3.

Tabel 3 – Korelasi variabel prediktor terhadap variabel respon

Variabel	Korelasi Pearson	Probability Value
X_1 vs Y	-0,271	0,449
X_2 vs Y	0,608	0,062
X_3 vs Y	0,723	0,018
X_4 vs Y	-0,397	0,256

Dari nilai korelasi pada Tabel 3 dapat dilihat bahwa masing-masing variabel prediktor memiliki korelasi yang cukup tinggi dengan variabel respon. Oleh karenanya, dapat disimpulkan bahwa variabel prediktor memiliki pengaruh terhadap variabel respon. Setelah melihat adanya korelasi maka dapat dibuat *scatter plot* masing-masing variabel prediktor terhadap variabel respon untuk mengetahui pola data. Berikut *scatter plot* masing-masing variabel prediktor dengan variabel respon.



Gambar 1 – Scatter plot variabel respon dengan variabel predictor

Dari Gambar 1 dapat dilihat bahwa variabel respon (Y) tidak dapat diasumsikan mengikuti pola hubungan tertentu dengan masing-masing variabel predictor (X). Oleh karenanya, data dimodelkan dengan pendekatan regresi nonparametrik kernel, dalam hal ini fungsi kernel yang digunakan yaitu Gaussian dengan estimator kernel Nadaraya Watson.

Untuk memodelkan data dengan regresi nonparametrik kernel diawali dengan pemilihan *bandwidth* optimal menggunakan algoritma *simulated annealing*. Algoritma ini melakukan pencarian lokal dengan hasil secara optimal, dengan kemudahan implementasi, sifat konvergensi, dan penggunaannya. Algoritma ini memanfaatkan teori peluang dalam penentuan global minimum dalam masalah optimasi (Noviardianto, dkk., 2019).

Dalam hal ini, algoritma ini digunakan untuk membantu meminimumkan nilai *Generalized Cross Validation* (GCV) guna menentukan *bandwidth* yang

optimal. Dari hasil percobaan, didapatkan nilai *bandwidth* optimal dan GCV minimum dengan *simulated annealing algorithm* seperti pada tabel 4.

Tabel 4 – Nilai *Bandwith* dan GCV Minimum

Keterangan	h_1	h_2	h_3	h_4
<i>Bandwidth</i> Optimal	0.5554 2837	1.2904 2927	0.9470 6041	0.9227 8896
GCV Minimum	0.000000000432613511			

Setelah diperoleh nilai *bandwidth* optimal maka diperoleh estimator Nadaraya Watson sebagai berikut, dengan nilai h_j tertera pada Tabel 4.

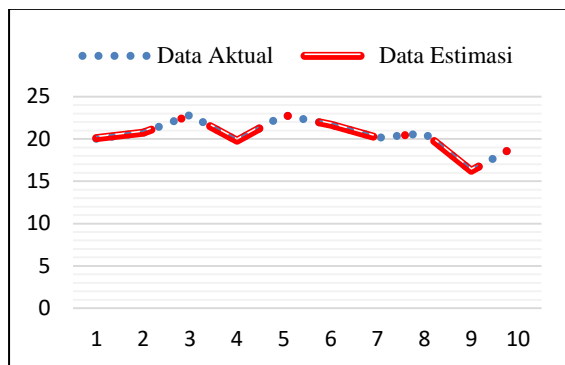
$$m(x_i) = \frac{\sum_{i=1}^n \prod_{j=1}^4 K\left(\frac{x_{ij}-x_j}{h_j}\right) y_i}{\sum_{i=1}^n \prod_{j=1}^4 K\left(\frac{x_{ij}-x_j}{h_j}\right)}$$

Adapun nilai estimasi data dengan bandwidth optimal disajikan pada Tabel 5.

Tabel 5 – Hasil estimasi data aktual y_i

No	Kota	Data Aktual	Data Estimasi
1	Lombok Barat	20,03	20,1009
2	Lombok Tengah	20,76	20,7237
3	Lombok Timur	22,83	22,7935
4	Sumbawa	19,80	19,8000
5	Dompu	22,81	22,8099
6	Bima	21,69	21,6900
7	Sumbawa Barat	20,14	20,1399
8	Lombok Utara	20,67	20,6699
9	Kota Mataram	16,23	16,2300
10	Kota Bima	19,32	19,3200

Berdasarkan Tabel 5, terlihat bahwa data hasil estimasi dengan menggunakan model yang diperoleh tidak berbeda jauh atau hampir sama dengan data aktualnya. Kota dengan jumlah angka kelahiran kasar tertinggi tahun 2018 adalah Lombok Timur. Adapun kurva regresi data estimasi dan data aktual dapat dilihat pada Gambar 2.



Gambar 2 – Kurva data estimasi dan data aktual

Untuk mengukur kelayakan model yang telah diperoleh digunakan koefisien determinasi (R^2) dan diperoleh nilai sebesar 0,992393265 atau 99,23%. Nilai itu sangat tinggi, artinya variabel prediktor mampu menjelaskan variabel respon sebesar 99,23% sisanya mampu dijelaskan oleh variabel yang lain. Setelah itu, dilakukan pengujian asumsi residual yaitu uji identik dengan membuat *scatter plot* antara

residual dengan variabel dugaan respon (\hat{y}) dan uji independensi residual dengan ACF residual. Hasil uji asumsi didapatkan sebagai berikut.

Dari Gambar 3 dapat dilihat bahwa *scatter plot* antara residual dengan variabel dugaan respon (\hat{y}) acak dan tidak menunjukkan suatu pola yang membentuk corong atau garis melengkung, maka dapat disimpulkan residual bersifat identik (Draper dan Smith, 1998). Selanjutnya, pada plot ACF residual dapat dilihat bahwa tidak ada *lag* yang memotong selang kepercayaan sehingga dapat disimpulkan residual bersifat independen. Jadi, asumsi identik dan independensi residual terpenuhi. Setelah itu, dilakukan perhitungan nilai ketepatan model dengan *Mean Absolute Percentage Error* (MAPE) dan diperoleh hasilnya sebesar 0,0007049%. Artinya model memiliki ketepatan paramalan yang tinggi.

4. Kesimpulan

Dari penelitian yang dilakukan diperoleh kesimpulan bahwa model yang diperoleh memiliki ketepatan yang tinggi dan mampu memodelkan variabel angka kelahiran kasar dengan rata-rata usia kawin pertama perempuan 10 tahun ke atas (X_1), persentase penduduk usia 10 tahun keatas yang sudah menikah (X_2), persentase perempuan usia 15 tahun yang tidak atau belum tamat SD (X_3), dan persentase perempuan usia 15 tahun ke atas yang tamat universitas atau sarjana (X_4) terhadap variabel angka kelahiran kasar (Y) di 10 kota/kabupaten di Provinsi Nusa Tenggara Barat dengan baik. Hal itu dapat dilihat dengan memperhatikan nilai koefisien determinasi yang relatif tinggi dan MAPE yang relatif rendah yang diperoleh.

DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). (2011). *Fertilitas Penduduk Indonesia: Hasil Sensus Penduduk 2010*.
- Budiantara, I. N., Ratnasari, V., Ratna, M., and Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Applied Mathematical Sciences*, 9(122), pp. 6083–6094.
- Chodijah, R. (2011). *Perilaku Fertilitas di Berbagai Kabupaten di Sumatera Selatan*. Jakarta: Brief Policy

- Dinas Kesehatan. (2016). *Profil Kesehatan Provinsi Nusa Tenggara Barat*. Nusa Tenggara Barat: Dinas Kesehatan.
- Draper, N. R., and Smith, H. (1998). *Applied Regression Analysis* (Third). Canada: John Wiley & Sons.
- Fardlyana, E. (2009). Pernikahan Usia Dini dan Permasalahannya. *Jurnal Sari Pediatri*, 2(1), pp. 136–141.
- Hardle, W. (1994). *Applied Nonparametric Regression*. Berlin: Humboldt-Universität zu Press.
- Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Models: An Introduction*. Berlin: Humboldt-Universität zu Press.
- Jones, G. W. dan Gubhaju, B. (2008). Trend at Age marriage in the Provinces of Indonesia. *Asia Research Institute Working Paper*, 105.
- Noviardianto, G. E., Novel, M., dan Legowo, M. B. (2019). Penggunaan Metode Simulated Annealing untuk Optimasi Penempatan Posisi Access Point pada Jaringan WI-FI. *Jurnal AL-AZHAR INDONESIA SERI SAINS DAN TEKNOLOGI*, 5(1), pp. 10–18.