

PEMODELAN ANGKA MORBIDITAS DENGAN MENGGUNAKAN REGRESI NONPARAMETRIK SPLINE DI INDONESIA

Megawati*, Bahriidin Abapihi, Irma Yahya, Ruslan, Lilis Laome, Mukhsar
Program Studi S1 Statistika FMIPA Universitas Halu Oleo
*Email: megawatilamu99@gmail.com

ABSTRACT

Morbidity is a condition where a person is said to be sick if the perceived health complaints cause disruption of daily activities, namely unable to carry out work activities, take care of the household, and normal activities as usual. The higher the morbidity, the worse the health status of the population. Therefore, a research is needed to find the best model so that it is able to explain related phenomena using an approach that is in accordance with the form of the data. One method that can be used is nonparametric spline regression. A spline is a segmented polynomial that has a common fusion point that shows changes in the behavior of the curve. The purpose of this study was to determine the modeling of morbidity rates using the Spline Nonparametric Regression Method in Indonesia. Based on the best model of morbidity rate in Indonesia, it is at three knot points with a minimum GCV (Generalized Cross-Validation) value of 4.97 and R^2 is 0.62 which means the ability of the response variable to explain the dependent variable is 62% while the rest is influenced by other variables outside the study.

Keywords: *Morbidity rate, Spline Nonparametric Regression*

ABSTRAK

Morbidity adalah kondisi seseorang dikatakan sakit apabila keluhan kesehatan yang dirasakan menyebabkan terganggunya aktivitas sehari-hari yaitu tidak dapat melakukan kegiatan bekerja, mengurus rumah tangga, dan kegiatan normal sebagaimana biasanya. Semakin tinggi morbidity, menunjukkan derajat kesehatan penduduk yang semakin buruk. Oleh karena itu dibutuhkan suatu penelitian untuk mencari model terbaik sehingga mampu menjelaskan fenomena terkait menggunakan pendekatan yang sesuai dengan bentuk data. salah satu metode yang dapat digunakan adalah Regresi nonparametrik spline. Spline merupakan potongan polinomial tersegmen yang memiliki titik perpaduan bersama yang menunjukkan terjadinya perubahan-perubahan perilaku kurva. Tujuan dari penelitian ini adalah untuk mengetahui pemodelan angka morbidity dengan menggunakan Metode Regresi Nonparametrik Spline di Indonesia. Berdasarkan model terbaik dari Angka Morbidity di Indonesia adalah pada tiga titik knot dengan nilai GCV (*Generalized Cross-Validation*) minimum 4.97 serta R^2 adalah 0.62 yang berarti kemampuan variabel respon dalam menjelaskan variabel terikatnya sebesar 62% sedangkan sisanya dipengaruhi oleh variabel lain diluar penelitian.

Kata Kunci: *Angka Morbidity, Regresi Nonparametrik Spline*

PENDAHULUAN

Morbidity (kesakitan) merupakan salah satu indikator yang digunakan untuk mengukur derajat kesehatan penduduk. Semakin tinggi morbidity, menunjukkan derajat kesehatan penduduk semakin buruk. Sebaliknya semakin rendah morbidity (kesakitan) menunjukkan derajat kesehatan penduduk yang semakin baik. Pengertian morbidity (kesakitan) adalah kondisi seseorang dikatakan sakit apabila keluhan kesehatan yang dirasakan mengganggu aktivitas sehari-hari yaitu tidak dapat melakukan kegiatan seperti bekerja, mengurus rumah tangga dan kegiatan lainnya secara normal sebagaimana biasanya (Hanun & Purhadi, 2013).

Analisis regresi adalah kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan dengan satu atau dua variabel yang menerangkan. Variabel yang diterangkan selanjutnya disebut sebagai variabel respon, sedangkan variabel yang menerangkan biasa disebut variabel bebas (Syilfi, 2012). Regresi nonparametrik adalah salah satu metode yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor dimana fungsi dari kurva regresi tidak diketahui. Pendekatan

nonparametrik digunakan untuk mengestimasi kurva regresi karena model tidak ditentukan terlebih dahulu seperti pada regresi parametrik. Dalam pandangan regresi nonparametrik data diharapkan mencari sendiri kurva regresi tanpa dipengaruhi oleh faktor subyektifitas peneliti (Litawati, 2013). Regresi spline merupakan potongan polinomial memainkan peranan penting dalam teori statistika. Potongan polinomial mempunyai sifat fleksibel dan efektif untuk mengenai sifat lokal atau data.

Berdasarkan penelitian yang dilakukan oleh Wulandari (2017) di Jawa Timur dengan menggunakan metode Regresi Nonparametrik Spline yang menghasilkan faktor-faktor yang mempengaruhi angka morbiditas di Jawa Timur adalah kepadatan penduduk, rata-rata lama sekolah, persentase penduduk miskin, Upah Minimum Kabupaten (UMK), persentase rumah tangga *Open Defecation* (OD), dan persentase rumah tangga dengan jarak sumber air minum ke tempat penampungan kotoran lebih dari 10 meter. Tujuan dari penelitian ini adalah untuk mengetahui pemodelan angka morbiditas dengan menggunakan Metode Regresi Nonparametrik Spline di Indonesia.

METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari aplikasi Allstats BPS. Data tersebut terdiri atas 3 variabel prediktor dan 1 variabel respon. Unit penelitian yang digunakan adalah seluruh Provinsi di Indonesia yaitu sebanyak 34 Provinsi pada tahun 2020.

Tabel 1. Definisi Operasional Variabel Penelitian

Variabel		Definisi Operasional	Satuan
Respon			
Angka Morbiditas	Y	Gangguan terhadap kondisi fisik maupun jiwa, termasuk karena kecelakaan, atau hal lain yang menyebabkan terganggunya kegiatan sehari-hari.	Persen
Prediktor			
Rata-rata Lama Sekolah	X_1	Jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal.	Persen
Persentase Penduduk Miskin	X_2	Penduduk yang memiliki rata-rata pengeluaran per kapita per bulan dibawah garis kemiskinan.	Persen
Air Minum Layak	X_3	Air minum yang terlindung meliputi air ledeng (keran), keran umum, hydrant umum, terminal air, penampungan air hujan (PAH) atau mata air dan sumur terlindung, sumur bor atau sumur pompa, yang jaraknya minimal 10 m dari pembuangan kotoran, penampungan limbah dan pembuangan sampah.	Persen

Prosedur penelitian ini meliputi langkah-langkah sebagai berikut:

1. Membuat analisis deskriptif terhadap karakteristik Angka Morbiditas di Indonesia.
2. Membuat *scatterplot* antara Angka Morbiditas (Y) dengan masing-masing variabel prediktor (X).
3. Memilih titik knot optimal berdasarkan GCV (*Generalized Cross-Validation*) minimum.
4. Melakukan pengujian signifikansi parameter dan pengujian asumsi residual model spline.
5. Membuat kesimpulan.

HASIL DAN PEMBAHASAN

Penulis akan membahas mengenai data angka morbiditas dan faktor-faktor yang diduga mempengaruhinya. Kemudian akan dilakukan analisis dengan menggunakan metode statistika deskriptif untuk mengetahui karakteristik dari variabel respon dan variabel-variabel prediktornya. Setelah itu, dilakukan pemodelan angka morbiditas di Indonesia pada tahun 2020 menggunakan metode regresi nonparametrik *spline*.

Deskriptif Data

Deskriptif adalah suatu metode statistika yang digunakan untuk dapat memperoleh informasi dari data yang merupakan hasil pengamatannya. Berikut deskriptif data variabel respon dan variabel prediktor yang digunakan dalam Tabel 3.1.

Tabel 2. Deskriptif Data Morbiditas dan Faktor yang diduga Berpengaruh

Variabel	Rata-rata	Minimum	Maksimum
Y	13.28	7.81	21.09
X ₁	8.649	6.690	11.130
X ₂	2.014	0.610	6.900
X ₃	85.41	62.47	99.84

Berdasarkan Tabel 2 diketahui bahwa variabel respon (Y) yaitu morbiditas memiliki rata-rata sebesar 13,28. Angka Morbiditas terendah adalah 7,81 berada di Kepulauan Riau, dan Angka Morbiditas tertinggi adalah 21,09 berada di Nusa Tenggara Barat. Rata-rata lama sekolah (X₁) memiliki nilai rata-rata adalah 8,649. Nilai terendah Rata-rata Lama Sekolah adalah 6,69 berada di Papua. Nilai tertinggi Rata-rata Lama Sekolah adalah 11,130 berada di DKI Jakarta. Persentase Penduduk Miskin (X₂) memiliki nilai rata-rata yaitu 2.014. Nilai terendah Persentase Penduduk Miskin adalah 0.61 berada di Bali. Nilai tertinggi Persentase Penduduk Miskin adalah 6.9 berada di Papua. Air Minum Layak (X₃) memiliki nilai rata-rata yaitu 85,41. Nilai terendah Air Minum Layak adalah 62,47 berada di Bengkulu. Nilai tertinggi Air Minum Layak adalah 99,84 berada di DKI Jakarta.

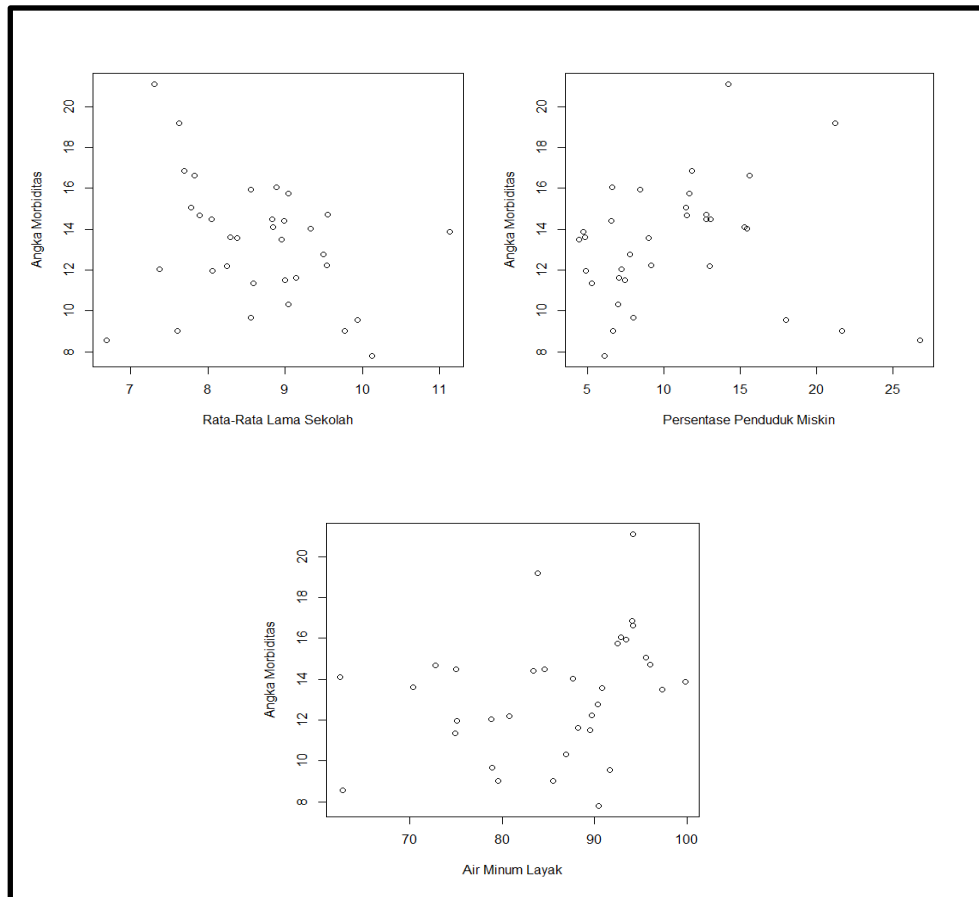
Scatterplot Morbiditas dan faktor-faktor yang mempengaruhinya

Mengetahui hubungan antara Morbiditas dengan faktor-faktor yang diduga mempengaruhi merupakan langkah awal dalam melakukan pemodelan dengan menggunakan pendekatan regresi nonparametrik *spline*. Untuk mengetahui bagaimana hubungan secara visual antara Morbiditas dengan faktor-faktor yang diduga mempengaruhi adalah dengan menggunakan *Scatterplot*. Berikut merupakan identifikasi hubungan antara Morbiditas dengan faktor-faktor yang diduga mempengaruhi.

Gambar 1 menunjukkan bahwa *Scatterplot* antara morbiditas dengan variabel-variabel yang diduga berpengaruh. Tidak menunjukkan adanya kecenderungan membentuk pola tertentu. Sehingga metode regresi nonparametrik *spline* cocok digunakan untuk pemodelan. Karena keseluruhan plot tidak membentuk pola tertentu, maka regresi nonparametrik *spline* dapat digunakan dalam pemodelan.

Pemilihan Model Terbaik

Titik knot optimum untuk satu knot, dua knot, dan tiga knot didapatkan, maka langkah selanjutnya adalah membandingkan nilai GCV (*Generalized Cross-Validation*) dari masing masing nilai knot optimum yang telah didapatkan. Berikut merupakan perbandingan GCV (*Generalized Cross-Validation*) minimum dan R^2 dengan berbagai titik knot yang ditunjukkan pada Tabel 3.



Gambar 1. Scatterplot morbiditas dengan variabel-variabel prediktor

Tabel 3. Perbandingan Nilai GCV (*Generalized Cross-Validation*) minimum dan R^2

Knot	GCV	R^2
Satu Knot	5.88	33%
Dua Knot	5.46	50%
Tiga Knot	4.97	62%

Dari Tabel 3 menunjukkan nilai GCV (*Generalized Cross-Validation*) yang dihasilkan dengan berbagai titik knot, dapat dilihat bahwa nilai GCV (*Generalized Cross-Validation*) yang paling minimum, adalah dihasilkan oleh pemilihan tiga titik knot sehingga tiga titik knot merupakan pemodelan terbaik. Semakin tinggi R^2 yang dihasilkan suatu model, maka semakin baik pula variabel-variabel prediktor dalam model tersebut dalam menjelaskan variabilitas variabel respon. Berikut adalah model terbaik tiga titik knot :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 (X_1 - k_1)^1 + \beta_3 (X_1 - k_2)^1 + \beta_4 (X_1 - k_3)^1 + \beta_5 X_2 + \beta_6 (X_2 - k_4)^1 + \beta_7 (X_2 - k_5)^1 + \beta_8 (X_2 - k_6)^1 + \beta_9 X_3 + \beta_{10} (X_3 - k_7)^1 + \beta_{11} (X_3 - k_8)^1 + \beta_{12} (X_3 - k_9)^1$$

$$y = 28.80 - 2.53X_1 + 8.87(X_1 - 9.04)^1 - 11.14(X_1 - 9.55)^1 + 12.18(X_1 - 10.12)^1 - 0.03X_2 + 2.14(X_2 - 2.26)^1 - 1.26(X_2 - 2.74)^1 - 3.94(X_2 - 3.76)^1 + 0.05X_3 - 1.29(X_3 - 87.66)^1 + 1.73(X_3 - 90.41)^1 - 1.11(X_3 - 94.13)^1$$

Selanjutnya akan di uji residual pada model terbaik.

Pengujian Residual

Pengujian asumsi residual dilakukan untuk mengetahui kelayakan suatu model regresi. Analisis regresi nonparametrik *spline* memiliki asumsi Identik, Independen dan berdistribusi Normal (IIDN). Apabila model regresi tidak memenuhi asumsi tersebut maka model tersebut tidak layak digunakan untuk memodelkan variabel respon.

Uji Identik

Asumsi residual identik terpenuhi apabila tidak ada indikasi terjadinya heteroskedastisitas, dengan kata lain varians residual harus homogen. Pengujian asumsi residual identik dilakukan menggunakan Uji *Bruesch Pagan* hipotesis sebagai berikut:

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

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

Hasil dari pengujian asumsi residual identik untuk model terbaik Uji *Bruesch Pagan* disajikan dengan menggunakan *software* R. Hasil Uji *Bruesch Pagan* akan ditampilkan pada Tabel berikut:

Tabel 4. Hasil Uji *Bruesch Pagan*

BP	Df	p-value
17.76	12	0.1232

Berdasarkan Tabel 4 diketahui bahwa nilai *p-value* sebesar 0,1232. Dengan taraf signifikansi α sebesar 5% maka didapatkan keputusan gagal tolak H_0 karena nilai *p-Value* $> \alpha$ (0,05), sehingga dapat disimpulkan bahwa tidak ada heteroskedastisitas atau dengan kata lain variansi antar residual sama. Hal ini berarti bahwa asumsi residual identik telah terpenuhi.

Uji Independen

Asumsi residual kedua yang harus dipenuhi adalah harus memenuhi asumsi independen. Metode yang digunakan untuk menguji asumsi independen menggunakan uji *Durbin Watson*. Berikut merupakan hipotesis yang digunakan.

$$H_0 : \rho = 0 \text{ (tidak ada korelasi antar residual)}$$

$$H_1 : \rho \neq 0 \text{ (ada korelasi antar residual)}$$

Hasil dari pengujian asumsi residual untuk model terbaik independen uji *Dubrin*. Disajikan dalam Tabel berikut:

Tabel 5. Hasil Uji *Durbin Watson*

DW	<i>p-value</i>
2.4337	0,7778

Berdasarkan Tabel 5 diketahui hasil analisis dapat diketahui bahwa *p-value* untuk pengujian asumsi residual independet adalah 0.7778. Berdasarkan nilai *p-value* yang lebih besar dari α (0,05) maka kesimpulan yang dapat diambil adalah gagal tolak H_0 . Ini artinya tidak terjadi autokorelasi pada residul atau asumsi residual independen sudah terpenuhi.

Uji Normalitas

Pengujian asumsi selanjutnya yaitu uji asumsi residual berdistribusi normal. Pengujian ini dilakukan dengan menggunakan uji *Lilliefors* (*Kolmogorov-Smirnov*). Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \varepsilon = N(\mu, \sigma^2) \text{ (Residual berdistribusi normal)}$$

$$H_0 : \varepsilon \neq N(\mu, \sigma^2) \text{ (Residual tidak berdistribusi normal)}$$

Hasil dari pengujian asumsi normalitas untuk tiga titik uji *Lilliefors* disajikan dalam Tabel berikut:

Tabel 6. Hasil Uji *Lilliefors*

D	<i>p</i> – <i>value</i>
0.09054	0,683

Berdasarkan Tabel 6 diketahui hasil analisis dapat diketahui bahwa *p*-value untuk pengujian asumsi normalitas adalah 0.683. Berdasarkan nilai *p*-value yang lebih besar dari α (0,05) maka kesimpulan yang dapat diambil adalah gagal tolak H_0 . Ini artinya residual berdistribusi normal.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa pemodelan terbaik dari Angka Morbiditas di Indonesia adalah terdapat pada tiga titik knot dengan nilai GCV (*Generalized Cross-Validation*) minimum 4.97 serta R^2 adalah 0.62 yang berarti kemampuan variabel terikat dalam menjelaskan variabel bebas sebesar 62% sedangkan sisanya dipengaruhi oleh variabel lain diluar penelitian. Semakin tinggi R^2 yang dihasilkan suatu model, maka semakin baik pula variabel-variabel prediktor dalam model tersebut dalam menjelaskan variabilitas variabel respon. Dari hasil perbandingan nilai GCV (*Generalized Cross-Validation*) yang dihasilkan dengan berbagai titik knot, maka dapat disimpulkan bahwa nilai GCV (*Generalized Cross-Validation*) paling minimum yang dihasilkan yaitu tiga titik knot yang merupakan model terbaik.

DAFTAR PUSTAKA

- Ardyanto, F. 2018. *Estimasi Interval Parameter Komponen Parametrik dan Spline Dalam Model Regresi Semiparametrik Campuran Spline dan Kernel* [Tesis]. Surabaya: Institut Teknologi Sepuluh Nopember, Fakultas Matematika dan Ilmu Pengetahuan Alam.
- Bintariningrum, M, F., & Budiantara, I, N. 2014. Pemodelan Regresi Nonparametrik Spline Truncated dan Aplikasinya Pada Angka Kelahiran Kasar di Surabaya. *Jurnal Sains dan Seni POMITS*. Vol. 3, No. 1.
- BPS. 2020. Angka Morbiditas Indonesia 2020. Aplikasi Allstats, Indonesia.
- BPS. 2020. Rata-Rata Lama Sekolah. Aplikasi Allstats, Indonesia.
- BPS. 2020. Rumah Tangga Menurut Provinsi, Tipe Daerah dan Sumber Air Minum Layak. Aplikasi Allstats, Indonesia.
- BPS. 2020. Persentase Penduduk Miskin Menurut Provinsi. Aplikasi Allstats, Indonesia.