

Analisis Faktor-Faktor yang Memengaruhi Morbiditas di Jawa Timur menggunakan Regresi Nonparametrik Spline Truncated

Vivia Faustine Gunawan¹, Anisah Aunillah², M. Zidane Ramadhan³, Salsabilla Alya Putri I.⁴

¹162112133020 Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

²162112133030 Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

³162112133046 Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

⁴162112133048 Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

Abstract—Morbiditas merupakan angka penyakit dalam suatu populasi pada kurun waktu tertentu. Morbiditas dapat menjadi cerminan derajat kesehatan masyarakat sebagaimana sesuai dengan SDGs tujuan ke-3 yang berfokus pada kesehatan dan kesejahteraan. Provinsi Jawa Timur pada tahun 2022 memiliki angka morbiditas mencapai 14,26% yang menunjukkan bahwa derajat kesehatannya masih rendah. Dilakukan penelitian dengan memodelkan faktor-faktor yang diduga berpengaruh terhadap morbiditas di Jawa Timur, yang mana terdapat variabel respon dan enam variabel prediktor. Digunakan metode regresi nonparametrik *spline truncated* dikarenakan pola hubungan morbiditas dan faktor-faktornya yang acak. Variabel yang digunakan, yaitu kepadatan penduduk, persentase penduduk miskin, rata-rata lama sekolah, upah minimum kabupaten/kota, dan persentase rumah tangga ber-PHBS berpengaruh signifikan terhadap morbiditas di Jawa Timur. Hasil penelitian menunjukkan bahwa model regresi nonparametrik *spline truncated* terbaik adalah menggunakan kombinasi titik knot 3, 3, 2, 3 dengan nilai GCV minimum sebesar 0,0254. Nilai koefisien determinasi dari model sebesar 89,7%.

Keywords—morbiditas, Jawa Timur, regresi nonparametrik, *spline truncated*, titik knot.

I. PENDAHULUAN

Sustainable Development Goals (SDGs) atau Tujuan Pembangunan Berkelanjutan adalah serangkaian target yang dirancang oleh Perserikatan Bangsa-Bangsa (PBB) untuk mengatasi berbagai tantangan global dan meningkatkan kesejahteraan manusia di muka bumi. SDGs berisikan 17 tujuan dan 169 target dan salah satunya adalah tujuan 3 yang berfokus pada kesehatan dan kesejahteraan. Derajat kesehatan masyarakat menjadi salah satu penanda keberhasilan pembangunan di sektor kesehatan, di mana tingkat derajat kesehatan mencerminkan kondisi kesehatan yang baik dalam suatu lingkungan. Semakin tinggi derajat kesehatan masyarakat, maka semakin sehat keadaan masyarakatnya. Situasi derajat kesehatan masyarakat salah satunya dapat

tercermin dari angka morbiditas [1].

Morbiditas (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 [2]. Morbiditas juga dapat diartikan sebagai angka kesakitan, baik insiden maupun prevalen dari suatu penyakit yang menggambarkan kejadian penyakit dalam suatu populasi pada kurun waktu tertentu [1]. Semakin tinggi angka kesakitan pada suatu wilayah, maka tingkat kesehatan masyarakat di wilayah tersebut semakin memburuk. Angka kesakitan mampu menggambarkan keadaan wilayah yang sebenarnya karena berkaitan erat dengan faktor lingkungan, faktor kesehatan, dan faktor sosial-ekonomi.

Pada tahun 2022, angka kesakitan (morbiditas) di Provinsi Jawa Timur mencapai 14,26%. Angka tersebut lebih tinggi dari angka kesakitan (morbiditas) nasional yang hanya mencapai 13,36% [3]. Hal tersebut menunjukkan bahwa penduduk di Provinsi Jawa Timur yang mengalami keluhan kesehatan dan angka tersebut mencerminkan jika derajat kesehatan di Provinsi Jawa Timur masih rendah.

Penelitian ini bertujuan untuk memodelkan faktor-faktor yang diduga memiliki pengaruh terhadap angka kesakitan atau morbiditas di Provinsi Jawa Timur. Adapun variabel-variabel prediktor yang digunakan pada penelitian ini mencakup persentase kepadatan penduduk, penduduk miskin, rata-rata lama sekolah, upah minimum kabupaten/kota, rumah tangga yang menerapkan Perilaku Hidup Bersih dan Sehat (PHBS), serta penduduk yang memiliki akses sanitasi layak. Metode yang digunakan pada penelitian ini adalah regresi nonparametrik *spline truncated* karena plot antara variabel prediktor dan variabel respon membentuk pola acak. *Spline truncated* adalah suatu fungsi yang memodifikasi fungsi polinomial menjadi bentuk potongan-potongan polinomial. *Spline truncated* memiliki beberapa kelebihan, yaitu dapat menangani karakteristik data yang memiliki perilaku berubah-ubah pada sub-interval tertentu, memiliki interpretasi statistik yang sederhana, dan memiliki penyajian visual yang baik [4]. Dengan dilakukannya penelitian ini, diharapkan pemerintah dapat mengupayakan program pembangunan kesehatan guna

II. TINJAUAN PUSTAKA

A. Morbiditas

Morbiditas dapat diartikan angka kesakitan, baik insiden maupun prevalen dari suatu penyakit yang menggambarkan kejadian penyakit dalam suatu populasi pada kurun waktu tertentu [1]. Angka kesakitan atau morbiditas dapat dihitung menggunakan rumus berikut.

$$AM = \frac{JPKK}{JP} \times 100$$

Dimana,

AM = Angka Morbiditas

$JPKK$ = Jumlah penduduk yang mengalami gangguan kesehatan dan terganggunya aktivitas

JP = Jumlah Penduduk

B. Statistika Deskriptif

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna [5]. Statistika deskriptif dapat disajikan secara visual ataupun dalam bentuk ukuran statistik. Statistika deskriptif secara visual dapat berupa tabel, grafik/diagram. Ukuran statistik seperti ukuran pemusatan data meliputi rata-rata (mean), median dan modus. Selain itu, terdapat ukuran penyebaran data berupa nilai minimum, maksimum, range dan varians.

C. MinMax Normalization

Rescaling data merupakan proses penskalaan data ke dalam interval tertentu yang ditetapkan [6]. Salah satu cara untuk *rescaling* data adalah dengan menggunakan metode *MinMax normalization*. *MinMax normalization* merupakan teknik normalisasi yang mengubah nilai data ke dalam rentang antara 0 dan 1 [7]. Persamaan untuk menghitung *MinMax normalization* dapat dihitung menggunakan rumus berikut.

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Dimana,

x_i : nilai tertentu yang akan dinormalisasi

x' : nilai hasil normalisasi

$\min(x)$: nilai minimal dari sebuah atribut

$\max(x)$: nilai maksimal dari sebuah atribut

D. Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon (dependen) dan prediktor (independen) yang tidak diketahui bentuk fungsinya, hanya diasumsikan fungsi *smooth* yang artinya memuat dalam suatu ruang fungsi tertentu, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi [8]. Berikut adalah model umum dari regresi nonparametrik.

$$y_i = m(t_i) + \varepsilon_i$$

Dimana,

y_i : Variabel respon pengamatan ke-i

t_i : Variabel prediktor pengamatan komponen nonparametrik ke-i

$m(t_i)$: Fungsi regresi yang tidak diketahui

ε_i : Residual pengamatan ke-i

E. Spline Truncated

Spline merupakan potongan polinomial yang memiliki sifat lebih fleksibel dibandingkan polinomial biasa sehingga mampu mengatasi pola data pada sub-interval tertentu dengan bantuan titik knot [9]. *Spline truncated* memiliki beberapa kelebihan, yaitu dapat menangani karakteristik data yang memiliki perilaku berubah-ubah pada sub-interval tertentu, memiliki interpretasi statistik yang sederhana, dan memiliki penyajian visual yang baik [4]. Fungsi *spline truncated* derajat p dengan titik-titik knot K_1, K_2, \dots, K_r dapat dituliskan dalam persamaan berikut.

$$g(x) = \sum_{j=0}^p \beta_j x^j + \sum_{k=0}^r \beta_{p+k} (x - K_k)_+^p$$

dimana $\sum_{j=0}^p \beta_j x^j$ merupakan fungsi *polinomial* dan $\sum_{k=0}^r \beta_{p+k} (x - K_k)_+^p$ merupakan fungsi *truncated* (potongan)

F. Pemilihan Titik Knot Optimal

Pemilihan titik knot digunakan untuk mencari estimator *spline truncated* yang paling sesuai. Untuk memperoleh model *spline truncated* terbaik bergantung pada pemilihan titik knot. Pemilihan titik knot optimal pada regresi *spline truncated* umumnya menggunakan metode *Generalized Cross-Validation* (GCV) [10]. Persamaan GCV didefinisikan sebagai berikut.

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

dimana,

$MSE(K_1, K_2, \dots, K_r) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, I adalah matriks identitas, dan $A(K_1, K_2, \dots, K_r) = X(X'X)^{-1}X'$.

G. Koefisien Determinasi

Koefisien determinasi (R^2) atau kuadrat koefisien korelasi merupakan proporsi atau persentase dari total variasi variabel respon yang dijelaskan oleh garis regresi. Koefisien determinasi ini digunakan untuk mengetahui persentase pengaruh yang terjadi antara variabel prediktor dan variabel respon. Semakin tinggi nilai koefisien determinasi maka semakin baik variabel prediktor dalam menjelaskan variabilitas variabel respon. Nilai koefisien determinasi dapat dihitung dengan rumus berikut.

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

H. Pengujian Parameter Model

Pengujian parameter model regresi dilakukan untuk mengetahui apakah variabel prediktor yang berpengaruh signifikan terhadap variabel respon atau tidak. Terdapat dua tahap pengujian parameter yaitu pengujian secara serentak dan secara parsial.

G.1. Uji Serentak

Hipotesis

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

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

Statistik Uji

$$F_{hitung} = \frac{\sum_{i=1}^n (\hat{y}_i - \underline{y})^2 / (p+r)}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - (p+r) - 1)}$$

Di mana p merupakan jumlah variabel prediktor kecuali β_0 dan r adalah jumlah titik knot. Daerah penolakan untuk uji serentak yaitu jika $F_{hitung} > F_{\alpha, (p+r, n-(p+r)-1)}$ atau $p\text{-value} < \alpha$

G.2. Uji Parsial

Hipotesis

$$H_0: \beta_j = 0$$

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

Statistik Uji

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

Daerah penolakan H_0 adalah apabila $|t_{hitung}| > t_{(\alpha/2, n-(p+r)-1)}$ atau $p\text{-value} < \alpha$.

I. Uji Asumsi Residual

Residual yang dihasilkan oleh model terbaik wajib memenuhi tiga asumsi yaitu identik, independen, dan berdistribusi normal.

I.1. Asumsi Identik

Asumsi identik merupakan asumsi yang mengharuskan varians residual sama atau identik. Apabila varians tidak sama maka akan terjadi heteroskedastisitas yang dapat mengakibatkan kerugian bagi efisiensi estimator. Salah satu cara untuk mendeteksi adanya heteroskedastisitas adalah dengan menggunakan uji *Glejser*.

Hipotesis :

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

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

di mana nilai v menunjukkan banyaknya parameter model *Glejser*. Daerah penolakannya yaitu tolak H_0 apabila $F_{hitung} > F_{\alpha; (v, n-v-1)}$ atau $p\text{-value} < \alpha$ yang dapat diartikan terdapat kasus heteroskedastisitas, sehingga asumsi residual identik tidak dapat terpenuhi.

I.2. Asumsi Independen

Asumsi independen merupakan asumsi yang mewajibkan antar residual tidak saling berkorelasi. Apabila terdapat korelasi antar residual maka akan terjadi autokorelasi. Salah satu uji yang digunakan untuk mendeteksi kasus autokorelasi yaitu uji *Durbin Watson*.

I.3. Asumsi Normalitas

Asumsi normalitas merupakan asumsi yang mewajibkan residual berdistribusi normal. Salah satu cara pengujian normalitas yaitu menggunakan uji *Kolmogorov-Smirnov*.

Hipotesis :

$$H_0: F_n(x) = F_0(x) \text{ (residual berdistribusi normal)}$$

$$H_1: F_n(x) \neq F_0(x) \text{ (residual tidak berdistribusi normal)}$$

Statistik uji :

$$D = \max |F_n(x) - F_0(x)|$$

Daerah penolakan yaitu tolak H_0 apabila $D > D_{(1-\alpha)}$ dengan nilai $D_{(1-\alpha)}$ adalah nilai kritis untuk uji *Kolmogorov Smirnov* satu sampel.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari berbagai sumber, diantaranya Laporan Statistik Kesejahteraan Rakyat tahun 2022, Badan Pusat Statistik Provinsi Jawa Timur, dan Profil Kesehatan Provinsi Jawa Timur tahun 2018. Data yang digunakan berisikan angka morbiditas beserta enam faktor yang diduga memiliki pengaruh. Sampel penelitian yang digunakan adalah 38 Kabupaten/Kota di Provinsi Jawa Timur pada tahun 2022.

B. Variabel Penelitian

Variabel yang digunakan pada penelitian ini disajikan pada Tabel 1.

Table 1. Variabel Penelitian

Variabel	Keterangan
Y	Angka Morbidity
x_1	Kepadatan penduduk
x_2	Persentase penduduk miskin
x_3	Rata-rata Lama Sekolah (RLS)
x_4	Upah Minimum Kabupaten/Kota (UMK)
x_5	Persentase rumah tangga ber-PHBS
x_6	Persentase penduduk dengan akses sanitasi layak

Berdasarkan Tabel 1 di atas, terdapat enam variabel prediktor, yaitu KP, PPM, RLS, UMK, sanitasi, dan PHBS serta variabel responnya adalah Angka Morbidity.

C. Langkah Analisis

Berikut adalah langkah-langkah analisis yang dilakukan pada penelitian ini dengan menggunakan bantuan *software RStudio* dan *Python*.

1. Mengumpulkan data dan mendeskripsikan data yang digunakan dalam penelitian melalui statistika deskriptif.
2. Melakukan data *preprocessing* yaitu normalisasi agar skala data berada pada rentang yang sama.
3. Mengidentifikasi pola hubungan masing-masing variabel prediktor dengan variabel respon melalui visualisasi *scatterplot*.
4. Memodelkan data menggunakan metode regresi nonparametrik *spline truncated* dengan 1 knot, 2 knot, 3 knot, dan kombinasi knot.
5. Memilih titik knot optimal berdasarkan nilai GCV terkecil.
6. Mendapatkan model regresi nonparametrik *spline truncated* terbaik melalui titik knot optimal.
7. Menghitung nilai koefisien determinasi R^2 .
8. Menguji parameter model yang telah didapatkan dengan uji serentak dan uji parsial.
9. Melakukan uji asumsi residual yang meliputi asumsi identik, asumsi independen, dan asumsi normalitas.
10. Menginterpretasikan model dan menarik kesimpulan.

IV. ANALISIS DAN PEMBAHASAN

A. Karakter Morbiditas di Jawa Timur dan Faktor-Faktor yang Diduga Mempengaruhinya

Berikut merupakan karakteristik morbiditas di Jawa Timur dan faktor-faktor yang diduga mempengaruhinya.

Table 2. Statistika Deskriptif Variabel Penelitian

Variabel	Mean	Variance	Min	Max
Y	14,56	22,23	5,01	29,24
x_1	2071,10	5509705,88	4141,00	8595,00
x_2	10,32	18,23	3,79	21,61
x_3	8,27	2,63	5,06	11,67
x_4	2478224,47	6,525E+11	1891567,00	4375479,00
x_5	82,16	152,58	51,64	96,41
x_6	51,92	243,84	18,20	83,20

Berdasarkan Tabel 2 diketahui bahwa variabel respon (y) yaitu morbiditas di Provinsi Jawa Timur memiliki nilai rata-rata sebesar 14,56 dengan varians data sebesar 22,23. Angka morbiditas tertinggi yaitu 29,24 yang terdapat di Ponorogo dan angka morbiditas terendah sebesar 5,01 berada pada Kota Surabaya.

B. Preprocessing Dataset

Preprocessing dilakukan dengan *rescaling* data berupa normalisasi terhadap data menggunakan metode *MinMax*. Normalisasi dilakukan untuk mengubah skala atau rentang variabel sehingga memudahkan dalam analisis data. Dengan mereskalakan variabel independen ke rentang yang serupa dapat memahami dampak relatif dari tiap-tiap variabel pada bentuk *spline* yang dihasilkan, dikarenakan *spline truncated* rentan terhadap variabel dengan skala besar yang mendominasi variabel lain.

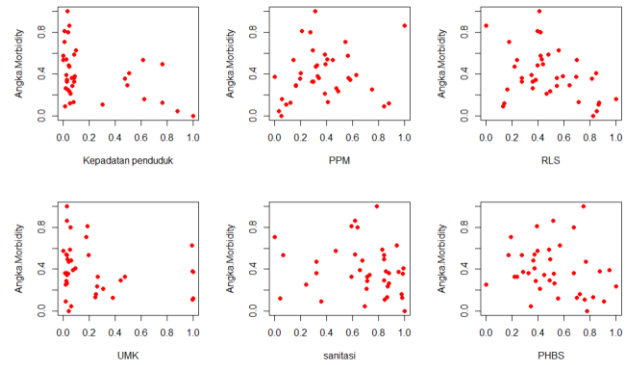
Table 3. Statistika Deskriptif Variabel Penelitian

Y	x_1	x_2	x_3	x_4	x_5	x_6
0.578	0	0.562	0.418	0	0.397	0.469
1	0.033	0.310	0.410	0.028	0.748	0.785
0.540	0.022	0.402	0.428	0.025	0.372	0.617
...
0.162	0.623	0.054	1	0.249	0.723	0.977
0	1	0.052	0.825	0.040	0.777	1
0.373	0.086	0	0.691	1	0.283	0.954

Ditampilkan hasil *rescaling* data menggunakan metode *MinMax normalization* untuk memperoleh data dengan skala yang sama pada Tabel 3. Hasil normalisasi data tersebut akan dimodelkan menggunakan metode regresi nonparametrik *spline truncated*.

C. Analisis Pola Hubungan Faktor-Faktor yang Diduga Mempengaruhi Morbiditas di Jawa Timur

Sebelum melakukan pemodelan, dilakukan identifikasi pola hubungan antara variabel respon dengan variabel prediktor menggunakan *scatterplot*.



Gambar 1. Scatterplot variabel Y dengan masing-masing variabel X

Berdasarkan Gambar 1 terlihat bahwa hubungan antara morbiditas dengan enam faktor yang diduga mempengaruhinya tidak membentuk pola tertentu, sehingga metode yang digunakan dalam penelitian ini yaitu regresi nonparametrik *spline truncated*.

D. Pemodelan Morbiditas di Jawa Timur

Pemodelan morbiditas di Jawa Timur menggunakan metode regresi nonparametrik *spline truncated* dengan satu, dua, tiga dan kombinasi knot. Pemilihan model terbaik didasari oleh nilai GCV yang paling minimum. Berikut merupakan nilai GCV minimum pada setiap pemodelan menggunakan satu, dua, tiga dan kombinasi titik knot.

Table 4. Perbandingan Nilai GCV

No	Knot	GCV Minimum
1	Satu titik knot	0.0561693
2	Dua titik knot	0.04474184
3	Tiga titik knot	0.0352729
4	Kombinasi knot (2,2,2,3,2,2)	0.02582217

Berdasarkan Tabel 4 menunjukkan bahwa nilai GCV yang paling minimum terdapat pada model *spline* dengan kombinasi titik knot 2, 2, 2, 3, 2, 2. Oleh karena itu, model *spline* dengan kombinasi knot merupakan model regresi nonparametrik *spline truncated* terbaik untuk pemodelan morbiditas di Jawa Timur. Berikut merupakan hasil estimasi parameter model regresi nonparametrik *spline truncated* terbaik menggunakan metode *Ordinary Least Square (OLS)*.

$$\hat{y} = 0.36288654 + 0.18765145X_1 + 0.09417232(X_1 - 0.4285714) - 0.71644331(X_1 - 0.8979592) - 0.43762287X_2 - 0.33387574(X_2 - 0.4285714) + 1.36875155(X_2 - 0.8979592) + 1.51022169X_3 + 0.43449441(X_3 - 0.4285714) + 0.05303436(X_3 - 0.8979592) + 0.21669502X_4 - 0.47162876(X_4 - 0.2040816) + 4.22348813(X_4 - 0.4081633) - 0.23135216(X_4 - 0.877551) - 0.68317156X_5 - 0.86990649(X_5 - 0.4285714) - 0.33168580(X_5 - 0.8979592) - 0.24182712X_6 + 0.20954791(X_6 - 0.4285714) - 0.29470358(X_6 - 0.8979592)$$

E. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter model dilakukan untuk mengetahui variabel prediktor apa saja yang signifikan berpengaruh terhadap morbiditas di Jawa Timur. Pengujian ini terdiri dari pengujian secara serentak dan parsial.

E.1. Pengujian Serentak

Hipotesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{19} = 0$$

$$H_1: \text{minimal terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, 19$$

Table 5. Analysis of Variance

Sumber Variasi	df	SS	MS	F_{hitung}	P- Value
Regresi	19	1,874	0,0986	8,0632	2,34342365302557e-05
Error	18	0,22	0,01220		
Total	37	2,094			

Pada tabel 5 menunjukkan bahwa nilai statistik uji F_{hitung} sebesar 8,0632 dengan p -value sebesar 2,34342365302557e-05. Nilai F_{hitung} (8,0632) lebih besar dari $F_{(0,05;19;18)}$ (2.20) dan p -value $< \alpha$ (0,05) sehingga keputusan yang didapat yaitu **Tolak H_0** yang artinya terdapat minimal satu variabel prediktor yang berpengaruh signifikan terhadap morbiditas di Jawa Timur. Oleh karena itu bisa dilanjutkan pengujian parameter secara parsial.

E.2. Pengujian Parsial

Hipotesis pada pengujian parsial adalah sebagai berikut

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, j = 1, 2, \dots, 19$$

Hasil pengujian parameter secara parsial ditampilkan pada Tabel 6.

Tabel 6. Hasil Pengujian Parameter secara Parsial

Variabel	Parameter	Estimator	P-value	Keputusan
konstan	β_0	0,34492	0,0794	Tidak Signifikan
	β_1	0,20813	0,1061	Tidak Signifikan
	β_2	0,08852	0,4366	Tidak Signifikan
X_1	β_3	-0,68134	0,0010	Signifikan
	β_4	-0,41950	0,0039	Signifikan
	β_5	-0,32980	0,0489	Signifikan
X_2	β_6	1,36940	0,0002	Signifikan
	β_7	1,49468	7,85e-05	Signifikan
	β_8	0,43495	0,1361	Tidak Signifikan
X_3	β_9	0,08816	0,7663	Tidak Signifikan
	β_{10}	0,21929	0,0146	Signifikan
	β_{11}	-0,43695	0,0004	Signifikan
X_4	β_{12}	4,34306	0,0019	Signifikan
	β_{13}	-0,22186	0,0348	Signifikan
	β_{14}	-0,66481	0,0005	Signifikan
X_5	β_{15}	-0,84787	0,0001	Signifikan
	β_{16}	-0,33739	0,0759	Tidak Signifikan

X_6	β_{17}	-0,24930	0,0661	Tidak Signifikan
	β_{18}	0,19184	0,0897	Tidak Signifikan
	β_{19}	-0,29320	0,0871	Tidak Signifikan

Berdasarkan Tabel 6 menunjukkan bahwa terdapat sembilan parameter yang tidak signifikan terhadap model karena nilai p -value $> \alpha$ (0,05) yaitu parameter $\beta_0, \beta_1, \beta_2, \beta_8, \beta_9, \beta_{16}, \beta_{17}, \beta_{18}$, dan β_{19} . Namun, kesebelas variabel prediktor berpengaruh signifikan terhadap morbiditas karena terdapat minimal satu parameter yang signifikan pada masing-masing variabel prediktor. Selain itu, pada variabel X_6 semua parameternya tidak signifikan sehingga perlu dihapus.

Table 7. Perbandingan Nilai GCV Baru setelah X_6 dihapus

No	Knot	GCV Minimum
1	Satu titik knot	0.04848857
2	Dua titik knot	0.03846718
3	Tiga titik knot	0.03023769
4	Kombinasi knot (3, 3, 3, 2, 3)	0.0253816

Berdasarkan Tabel 7 menunjukkan bahwa nilai GCV yang paling minimum terdapat pada model *spline* dengan kombinasi titik knot 3, 3, 3, 2, 3. Oleh karena itu, model *spline* dengan kombinasi knot merupakan model regresi nonparametrik *spline truncated* terbaik untuk pemodelan morbiditas di Jawa Timur. Berikut merupakan hasil estimasi parameter model regresi nonparametrik *spline truncated* terbaik menggunakan metode *Ordinary Least Square* (OLS).

$$\hat{y} = 2.12920855 - 0.65434357X_1 - 0.32277435(X_1 - 0.06122449) - 0.40700097(X_1 - 0.2857143) - 0.52971758(X_1 - 0.4693878) - 0.36519633X_2 + 0.15514701(X_2 - 0.06122449) - 0.50012044(X_2 - 0.2857143) + 0.42547728(X_2 - 0.4693878) - 1.77413448X_3 - 0.23694427(X_3 - 0.06122449) + 0.15916876(X_3 - 0.2857143) - 0.37350293(X_3 - 0.4693878) + 0.07911775X_4 - 0.31100157(X_4 - 0.2040816) - 0.04964405(X_4 - 0.6326531) - 0.59688306X_5 - 0.38194336(X_5 - 0.06122449) - 1.26591435(X_5 - 0.2857143) - 1.01234001(X_5 - 0.4693878)$$

Table 8. Analysis of Variance setelah X_6 dihapus

Sumber Variasi	df	SS	MS	F_{hitung}	P- Value
Regresi	19	1,878	0,0988	8,2196	2,03753547721271e-05
Error	18	0,216	0,012		
Total	37	2,094			

Pada tabel 8 menunjukkan bahwa nilai statistik uji F_{hitung} sebesar 8,2196 dengan p -value sebesar 2,03753547721271e-05. Nilai F_{hitung} (8,2196) lebih besar dari $F_{(0,05;19;18)}$ (2.20) dan p -value $< \alpha$ (0,05) sehingga keputusan yang didapat yaitu **Tolak H_0** yang artinya terdapat minimal satu variabel prediktor yang berpengaruh signifikan terhadap morbiditas di Jawa Timur. Oleh karena itu bisa dilanjutkan pengujian parameter secara parsial.

Tabel 9. Hasil Pengujian Parameter secara Parsial setelah X_6 dihapus

Variabel	Parameter	Estimator	P-value	Keputusan
konstan	β_0	2.12921	3,21e-07	Signifikan

X_1	β_1	-0.65434	5.10e-05	Signifikan
	β_2	-0.32277	0,166047	Tidak Signifikan
	β_3	-0.40700	0,017171	Signifikan
	β_4	-0.52972	0.001192	Signifikan
X_2	β_5	-0.36520	0.079825	Tidak Signifikan
	β_6	0.15515	0.266878	Tidak Signifikan
	β_7	-0.50012	0.009325	Signifikan
	β_8	0.42548	0.178079	Tidak Signifikan
X_3	β_9	-1.77413	3.64e-06	Signifikan
	β_{10}	-0.23694	0.400314	Tidak Signifikan
	β_{11}	0.15917	0.627589	Tidak Signifikan
	β_{12}	-0.37350	0.304632	Tidak Signifikan
X_4	β_{13}	0.07912	0.331209	Tidak Signifikan
	β_{14}	-0.31100	0.04936	Signifikan
	β_{15}	-0.04964	0.668738	Tidak Signifikan
X_5	β_{16}	-0.59688	0.003984	Signifikan
	β_{17}	-0.38194	0.032174	Signifikan
	β_{18}	-1.26591	0.000298	Signifikan
	β_{19}	-1.01234	0.000631	Signifikan

Berdasarkan Tabel 9 menunjukkan bahwa terdapat sembilan parameter yang tidak signifikan terhadap model karena nilai p -value $> \alpha$ (0,05) yaitu parameter $\beta_2, \beta_5, \beta_6, \beta_8, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$, dan β_{15} . Namun, kesebelas variabel prediktor berpengaruh signifikan terhadap morbiditas karena terdapat minimal satu parameter yang signifikan pada masing-masing variabel prediktor.

F. Pengujian Asumsi Residual

Residual hasil pemodelan menggunakan regresi nonparametrik *spline truncated* harus memenuhi asumsi residual identik, independen, dan berdistribusi normal. Berikut merupakan hasil pengujian asumsi residual:

F.1. Asumsi Identik

Asumsi identik (*homoskedastisitas*) merupakan asumsi yang mengharuskan varians residual sama/identik dan tidak terjadi kasus heteroskedastisitas. Berikut merupakan hasil pengujian asumsi identik dengan menggunakan uji *Glejser*.

Tabel 10. Hasil Pengujian Statistik Uji *Glejser*

Sumber	df	SS	MS	F_{hitung}	P- Value
Regresi	19	0,051	0,0027	0,8429	0,643106385210548
Error	18	0,058	0,0032		
Total	37	0,109			

Berdasarkan Tabel 10, diperoleh nilai statistik uji F_{hitung} sebesar 0,8429 dengan p -value sebesar 0,643106385210548. Nilai F_{hitung} (0,8429) lebih kecil dari $F_{(0,05;19;18)}$ (2,20) dan p -value $> \alpha$ (0,05) sehingga keputusan yang didapat yaitu **Gagal Tolak H_0** . Hal ini dapat diartikan bahwa residual telah memiliki varians yang sama/identik dan tidak terjadi kasus heteroskedastisitas sehingga residual telah memenuhi asumsi identik.

F.2. Asumsi Independen

Asumsi independen merupakan asumsi yang mengharuskan antar residual tidak saling berkorelasi. Apabila terdapat korelasi antar residual maka akan terjadi autokorelasi. Berikut merupakan hasil pengujian asumsi independen dengan uji *Durbin Watson*.

Tabel 11. Hasil Pengujian Statistik Uji *Durbin Watson*

DW	P-value
2,018333	0,632

Berdasarkan nilai p -value yang di dapat yaitu sebesar 0,632 $> \alpha$ (0,05), dapat disimpulkan **Gagal Tolak H_0** yang berarti tidak terjadi autokorelasi pada model, sehingga asumsi residual independen telah terpenuhi.

F.3. Asumsi Distribusi Normal

Asumsi normalitas merupakan asumsi yang mengharuskan residual dari model berdistribusi normal. Berikut merupakan hasil pengujian asumsi distribusi normal dengan uji *Kolmogorov-Smirnov*.

Tabel 12. Hasil Pengujian Statistik Uji *Kolmogorov-Smirnov*

D	P-value
0,1376584	0.4290387

Berdasarkan nilai p -value yang di dapat yaitu sebesar 0,4290387 $> \alpha$ (0,05), dapat disimpulkan **Gagal Tolak H_0** yang berarti residual berdistribusi normal dan asumsi normalitas residual telah terpenuhi.

G. Interpretasi Model Terbaik

Berdasarkan hasil analisis yang telah dilakukan sebelumnya, model regresi nonparametrik *spline* dengan kombinasi titik knot 3,3,3,2,3 telah memenuhi asumsi residual sehingga model tersebut merupakan model terbaik dalam memodelkan morbiditas di Jawa Timur. Nilai koefisien determinasi model sebesar 89,7 persen yang artinya keenam variabel prediktor mampu menjelaskan variabilitas morbiditas di Jawa Timur sebesar 89,7 persen sedangkan sisanya 10,3 persen dijelaskan oleh variabel lain diluar model. Berikut adalah interpretasi model pada setiap variabel yang signifikan terhadap morbiditas di Provinsi Jawa Timur.

1. Model *Spline* pada Variabel Kepadatan Penduduk (x_1)
Pengaruh variabel x_1 terhadap morbiditas di Jawa Timur apabila variabel lain selain variabel x_1 diasumsikan konstan adalah sebagai berikut.
 $\hat{y} = -0,654X_1 - 0,323(X_1 - 0,061) - 0,407(X_1 - 0,286) - 0,529(X_1 - 0,469)$

$$= \begin{cases} -0,654X_1, & X_1 < 0,061 \\ -0,977X_1 + 0,019, & 0,061 \leq X_1 < 0,286 \\ -1,384X_1 + 0,116, & 0,286 \leq X_1 < 0,469 \\ -1,913X_1 + 0,248, & X_1 \geq 0,469 \end{cases}$$

Berdasarkan persamaan model diatas, maka diketahui apabila kepadatan penduduk suatu kabupaten/kota di Jawa Timur kurang dari 0,061 dan mengalami pertambahan satu satuan maka morbiditas mengalami penurunan sebesar 0,654 satuan. Apabila kabupaten/kota memiliki kepadatan penduduk pada interval 0,061 hingga 0,286 dan bertambah satu satuan maka morbiditas akan mengalami penurunan sebesar 0,977 satuan. Apabila kabupaten/kota memiliki kepadatan penduduk pada interval 0,286 hingga 0,469 dan bertambah satu satuan maka morbiditas mengalami penurunan sebesar -1,384 satuan. Dan Apabila kabupaten/kota memiliki kepadatan penduduk lebih dari 0,469 dan bertambah satu satuan maka morbiditas akan mengalami penurunan sebesar 1,913 satuan.

2. Model *Spline* pada Variabel Persentase Penduduk Miskin (x_2)

Pengaruh variabel x_2 terhadap morbiditas di Jawa Timur apabila variabel lain selain variabel x_2 diasumsikan konstan adalah sebagai berikut.

$$\hat{y} = -0,362 + 0,155(X_2 - 0,061) - 0,500(X_2 - 0,286) + 0,425(X_2 - 0,469)$$

$$= \begin{cases} -0,362X_2, & X_2 < 0,061 \\ -0,207X_2 - 0,009, & 0,061 \leq X_2 < 0,286 \\ -0,707X_2 + 0,143, & 0,286 \leq X_2 < 0,469 \\ -0,282X_2 - 0,199, & X_2 \geq 0,469 \end{cases}$$

Berdasarkan model tersebut dapat diketahui apabila kabupaten/kota memiliki persentase penduduk miskin kurang dari 0,061 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 0,362 satuan. Apabila kabupaten/kota memiliki persentase penduduk miskin pada interval 0,061 hingga 0,286 dan naik satu satuan maka morbiditas akan turun sebesar 0,207 satuan. Apabila kabupaten/kota memiliki persentase penduduk miskin pada interval 0,286 hingga 0,469 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 0,707 satuan. Dan apabila kabupaten/kota memiliki persentase penduduk miskin lebih dari 0,469 dan naik satu satuan maka morbiditas akan naik sebesar 0,282 satuan.

3. Model *Spline* pada Variabel Rata-Rata Lama Sekolah (x_3)

Pengaruh variabel x_3 terhadap morbiditas di Jawa Timur apabila variabel lain selain variabel x_3 diasumsikan konstan adalah sebagai berikut.

$$\hat{y} = -1,774X_3 - 0,236(X_3 - 0,061) + 0,159(X_3 - 0,286) - 0,374(X_3 - 0,469)$$

$$= \begin{cases} -1,774X_3, & X_3 < 0,061 \\ -2,01X_3 + 0,014, & 0,061 \leq X_3 < 0,286 \\ -1,851X_3 - 0,045, & 0,286 \leq X_3 < 0,469 \\ -2,225X_3 + 0,175, & X_3 \geq 0,469 \end{cases}$$

Berdasarkan model tersebut dapat diketahui apabila rata-rata lama sekolah kabupaten/kota kurang dari 0,061 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 1,774 satuan. Apabila suatu kabupaten/kota memiliki rata-rata lama sekolah pada interval 0,061 hingga 0,286 dan mengalami kenaikan satu satuan maka morbiditas akan mengalami penurunan sebesar 2,01 satuan. Apabila suatu kabupaten/kota memiliki rata-rata lama sekolah pada interval 0,286 hingga 0,469 dan bertambah satu satuan maka akan mengalami penurunan sebesar 1,851 akan tetapi nilai tersebut sedikit naik jika dibandingkan dengan interval sebelumnya. Dan apabila suatu kabupaten/kota memiliki rata-rata lama sekolah lebih dari 0,469 dan mengalami kenaikan satu satuan maka morbiditas akan mengalami penurunan sebesar 2,225 persen.

4. Model *Spline* pada Variabel Upah Minimum Kabupaten/Kota (x_4)

Pengaruh variabel x_4 terhadap morbiditas di Jawa Timur apabila variabel lain selain variabel x_4 diasumsikan konstan adalah sebagai berikut.

$$\hat{y} = 0,079X_4 - 0,311(X_4 - 0,204) - 0,049(X_4 - 0,633)$$

$$= \begin{cases} 0,079X_4, & X_4 < 0,204 \\ -0,232X_4 + 0,063, & 0,204 \leq X_4 < 0,633 \\ -0,281X_4 + 0,031, & X_4 \geq 0,633 \end{cases}$$

Berdasarkan model tersebut dapat diketahui bahwa apabila kabupaten/kota memiliki UMK kurang dari 0,204 dan naik satu satuan maka morbiditas akan mengalami kenaikan sebesar 0,079 satuan. Apabila kabupaten/kota memiliki UMK pada interval 0,204 hingga 0,633 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 0,232 satuan. Dan apabila kabupaten/kota memiliki UMK lebih dari 0,633 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 0,281 satuan.

5. Model *Spline* pada Variabel Persentase Rumah Tangga Ber-PHBS (x_5)

Pengaruh variabel x_5 terhadap morbiditas di Jawa Timur apabila variabel lain selain variabel x_5 diasumsikan konstan adalah sebagai berikut.

$$\hat{y} = -0,596X_5 - 0,381(X_5 - 0,061) - 1,265(X_5 - 0,286) - 1,012(X_5 - 0,469)$$

$$= \begin{cases} -0,596X_5, & X_5 < 0,061 \\ -0,977X_5 + 0,023, & 0,061 \leq X_5 < 0,286 \\ -2,242X_5 + 0,362, & 0,286 \leq X_5 < 0,469 \\ -3,254X_5 + 0,475, & X_5 \geq 0,469 \end{cases}$$

Berdasarkan model tersebut dapat diketahui apabila persentase rumah tangga ber-PHBS di kabupaten/kota kurang dari 0,061 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 0,596 satuan. Apabila persentase rumah tangga ber-PHBS di kabupaten/kota berada pada interval 0,061 hingga 0,286 dan naik satu satuan maka morbiditas mengalami penurunan sebesar 0,977 satuan. Apabila persentase rumah tangga ber-PHBS di kabupaten/kota berada pada interval 0,286 hingga 0,469 dan naik satu

satuan maka morbiditas mengalami penurunan sebesar 2,242 satuan. Dan apabila persentase rumah tangga ber-PHBS di kabupaten/kota lebih dari 0,469 dan naik satu satuan maka morbiditas akan mengalami penurunan sebesar 3,254 satuan.

V. KESIMPULAN DAN SARAN

Angka morbiditas di Provinsi Jawa Timur menjadi perhatian dengan angka yang lebih tinggi dari rata-rata nasional. Morbiditas digambarkan sebagai indikator kesehatan yang mencerminkan kondisi kesehatan suatu populasi pada waktu tertentu. Dilakukan penelitian regresi nonparametrik *spline truncated* untuk memodelkan faktor-faktor yang mempengaruhi morbiditas di Provinsi Jawa Timur. Metode *spline truncated* digunakan karena pola hubungan antara variabel prediktor dan respon berpola acak. Dari penelitian yang dilakukan, diperoleh bahwa morbiditas di Provinsi Jawa Timur memiliki nilai rata-rata sebesar 14,56 dan varians data sebesar 22,23. Terdapat morbiditas tertinggi di Ponorogo sebesar 29,24, sedangkan morbiditas terendah di Kota Surabaya sebesar 5,01.

Data yang digunakan untuk pemodelan telah dilakukan normalisasi dengan metode *MinMax*. Setelah terpenuhi uji signifikansi parameter dan uji asumsi residual, didapatkan pemodelan regresi nonparametrik *spline truncated* terbaik untuk memodelkan morbiditas di Jawa Timur adalah model dengan kombinasi titik knot 3, 3, 3, 2, 3 dengan nilai GCV minimum sebesar 0,0254. Nilai koefisien determinasi model sebesar 89,7% dengan kelima variabel prediktor, yaitu kepadatan penduduk (x_1), persentase penduduk miskin (x_2), rata-rata lama sekolah (x_3), upah minimum kabupaten/kota (x_4), dan persentase rumah tangga ber-PHBS (x_5).

Saran untuk penelitian selanjutnya ialah dapat dilakukan penambahan jumlah variabel prediktor yang relevan dan cermat dengan mempertimbangkan pemilihan variabel agar dapat memperoleh model yang lebih optimal. Dari faktor-faktor yang diperoleh secara signifikan paling berpengaruh terhadap morbiditas di Jawa Timur, seperti persentase penduduk miskin, rata-rata lama sekolah, serta kepadatan penduduk, diharapkan pemerintah dapat memperhatikan maupun mempertimbangkan faktor tersebut. Sebaiknya, pemerintah dapat meningkatkan kualitas pendidikan, mengurangi tingkat kemiskinan, serta melakukan kampanye penyuluhan PHBS kepada masyarakat guna menurunkan angka morbiditas di wilayah tersebut.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia Tahun 2009*. Jakarta, 2010.
- [2] D. Hanum, D. D. H. Hanum, and P. P. Purhadi, "Faktor-faktor yang Mempengaruhi Morbiditas Penduduk Jawa Timur dengan Multivariate Geographically Weighted Regression (MGWR)," *Jurnal Sains dan Seni ITS*, vol. 2, no. 2, pp. D189–D194, Aug. 2013, Accessed: Dec. 18, 2023. [Online]. Available: http://ejurnal.its.ac.id/index.php/sains_seni/article/view/4957.
- [3] Direktorat Statistik Kesejahteraan Rakyat, *Statistik Kesejahteraan Rakyat 2022*. Badan Pusat Statistik, 2022.
- [4] A. T. R. Dani, N. Y. Adrianingsih, A. Ainurrochmah, and R. Sriningsih, "Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern," *JLE*, vol. 2, no. 1, pp. 37–43, May

- 2021, doi: 10.51402/jle.v2i1.30.
- [5] R. Walpole, *Introduction to Statistics*. New York: Macmillan Publishing Co. Inc, 1995.
- [6] B. Nugraha, "Regresi nonparametrik polinomial lokal untuk memodelkan inflasi di Indonesia," Jun. 2023.
- [7] I. Permana and F. N. S. Salisah, "Pengaruh normalisasi data terhadap performa hasil klasifikasi algoritma backpropagation," *IJIRSE*, vol. 2, no. 1, pp. 67–72, Mar. 2022, doi: 10.57152/ijirse.v2i1.311.
- [8] R. L. Eubank, *Nonparametric regression and spline smoothing*. CRC Press, 1999.
- [9] I. W. Rosanti and I. N. Budiantara, "Pemodelan Faktor-Faktor Yang Mempengaruhi Morbiditas Di Jawa Tengah Menggunakan Regresi Nonparametrik Spline Truncated," *Inferensi*, vol. 3, no. 2, pp. 107–114, Sep. 2020, Accessed: Dec. 18, 2023. [Online]. Available: <https://iptek.its.ac.id/index.php/inferensi/article/view/7712>.
- [10] L. P. S. Pratiwi, "Pemilihan Titik Knot Optimal Model Spline Truncated Dalam Regresi Nonparametrik Multivariabel dengan GCV," *JMAT*, vol. 10, no. 2, p. 78, Dec. 2020, doi: 10.24843/JMAT.2020.v10.i02.p125.