

Pemodelan Faktor-Faktor yang Mempengaruhi Persentase Kriminalitas di Jawa Timur dengan Pendekatan Regresi Semiparametrik Spline

Sherly Mega Tri Marina Dan I Nyoman Budiantara

Jurusan Statistika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: i_nyoman_b@statistika.its.ac.id

Abstrak—Kondisi ekonomi yang semakin tidak menentu dan suasana sosial yang tidak mendukung dapat menimbulkan masalah yang meresahkan masyarakat seperti tindak kriminal. Padahal, rasa aman adalah salah satu kebutuhan manusia yang harus terpenuhi. Menurut publikasi BPS berjudul Statistik Indonesia, pada tahun 2009 hingga tahun 2011 Provinsi Jawa Timur termasuk dalam lima besar Provinsi dengan angka tindak pidana tertinggi di Indonesia. Melihat betapa krusialnya masalah kriminalitas khususnya di Jawa Timur, pemodelan faktor-faktor yang mendorong adanya tindak kriminal sangat dibutuhkan. Pemodelan faktor-faktor yang mempengaruhi persentase kriminalitas ini dilakukan dengan menggunakan regresi semiparametrik Spline. Regresi Spline memiliki titik-titik knot dimana terdapat perubahan pola perilaku data. Titik knot dipilih berdasarkan nilai GCV minimum dan prinsip parsimoni. Berdasarkan penelitian yang telah dilakukan, terdapat enam variabel yang berpengaruh signifikan terhadap persentase kriminalitas di Jawa Timur. Variabel tersebut antara lain kepadatan penduduk, tingkat pengangguran terbuka, persentase penduduk miskin, persentase penduduk yang tidak pernah sekolah, persentase korban penyalahgunaan NAPZA, dan persentase keluarga bermasalah. Model tersebut dapat menjelaskan variabilitas persentase kriminalitas di Jawa Timur sebesar 92,13% dengan MSE sebesar 0,00018.

Kata Kunci— persentase kriminalitas, regresi semiparametrik spline, titik knot, GCV, parsimoni.

I. PENDAHULUAN

DEWASA ini, permasalahan yang dihadapi masyarakat semakin rumit. Kondisi ekonomi yang semakin tidak menentu dan suasana sosial yang tidak mendukung dapat menimbulkan masalah yang meresahkan masyarakat seperti tindak kriminal. Padahal rasa aman merupakan salah satu kebutuhan manusia (*safety needs*).

Menurut BPS [1], pada tahun 2009 Provinsi Jawa Timur menempati urutan tertinggi kedua angka tindak pidananya

setelah DKI Jakarta. Setelah itu hingga tahun 2011, Jawa Timur termasuk dalam lima besar Provinsi dengan angka tindak pidana tertinggi di Indonesia. Melihat fakta-fakta tersebut, maka analisis tentang faktor-faktor yang mempengaruhi tindak pidana akan sangat diperlukan.

Data kriminalitas dalam penelitian ini nantinya akan dimodelkan dengan menggunakan regresi Semiparametrik Spline. Metode tersebut dipilih karena berdasarkan plot pada studi awal yang telah dilakukan, diketahui bahwa sebagian variabel-variabel yang mempengaruhi persentase kriminalitas di Jawa Timur polanya diketahui bentuknya dan sebagian lagi tidak diketahui bentuk polanya. Spline adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen [2].

Penelitian sebelumnya tentang tingkat kriminalitas pernah dilakukan oleh Hakim [3] yang melakukan analisis determinan terhadap tingkat kejahatan properti di Jawa Tahun 2007. Selain itu terdapat pula Herdianto [4] yang menganalisis faktor-faktor yang mempengaruhi tingkat kriminalitas di Indonesia dari pendekatan ekonomi. Putra [5] juga melakukan identifikasi faktor-faktor yang mempengaruhi tindak kriminal berdasarkan karakteristik pelaku kriminal dengan metode pohon klasifikasi. Sementara itu terdapat beberapa penelitian sebelumnya yang juga menggunakan metode regresi spline, salah satunya yaitu Dwimayanti [6] yang mengaplikasikan regresi spline untuk memperkirakan tingkat fertilitas.

Secara umum terdapat dua tujuan dalam penelitian ini yaitu mendiskripsikan gambaran umum kriminalitas di Jawa Timur dan faktor-faktor yang diduga berpengaruh, serta Menentukan variabel-variabel yang berpengaruh terhadap persentase kriminalitas di Jawa Timur dengan pendekatan regresi semiparametrik Spline.

II. TINJAUAN PUSTAKA

A. Regresi Nonparametrik Spline

Pendekatan regresi nonparametrik dilakukan apabila pola data tidak mengikuti pola tertentu (linier, kuadrat, kubik,dll). Dengan demikian diharapkan data mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh subjektifitas

penelitinya. Dengan kata lain, regresi nonparametrik memiliki fleksibilitas yang tinggi [7].

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

dengan $f(t_i)$ merupakan kurva regresi yang dihipotesiskan dengan fungsi spline berorde p dengan titik knot k_1, k_2, \dots, k_r yang dapat diberikan oleh persamaan :

$$f(t_i) = \sum_{j=0}^p \beta_j t_i^j + \sum_{j=1}^r \beta_{p+j} (t_i - k_j)_+^p \quad (2)$$

Apabila persamaan (2.7) disubstitusikan kedalam persamaan (2.6) maka akan diperoleh persamaan regresi nonparametrik spline sebagai berikut.

$$y_i = \sum_{j=0}^p \beta_j t_i^j + \sum_{j=1}^r \beta_{p+j} (t_i - k_j)_+^p + \varepsilon_i \quad (3)$$

Fungsi $(t_i - k_j)_+^p$ merupakan fungsi potongan yang diberikan oleh

$$(t_i - k_j)_+^p = \begin{cases} (t_i - k_j)^p, & t_i \geq k_j \\ 0, & t_i < k_j \end{cases} \quad (4)$$

B. Regresi Semiparametrik

Apabila dalam suatu analisis regresi terdapat dua komponen yaitu komponen parametrik dan komponen nonparametrik, maka disebut regresi semiparametrik. Misalkan terdapat data berpasangan (x_i, y_i, t_i) , dan hubungan antara x_i, y_i , dan t_i diasumsikan mengikuti model regresi semiparametrik :

$$y_i = x_i' \beta + f(t_i) + \varepsilon_i, t_i \in [a, b], i = 1, 2, \dots, n \quad (5)$$

variabel prediktor $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ berhubungan secara parametrik dengan variabel respon y_i . Sementara itu, variabel prediktor t_i berhubungan secara nonparametrik dengan y_i . Jika kurva regresi f pada persamaan (5) dihipotesiskan dengan fungsi Spline (3) maka diperoleh regresi semiparametrik spline.

C. Pemilihan Titik Knot Optimal

Titik knot optimal dicari untuk mendapatkan model regresi spline terbaik yang paling sesuai dengan data. Menurut Wahba[8] dan Wang(1998) dalam Oktaviana[9], salah satu metode yang banyak dipakai dan disukai dalam memilih titik knot optimal adalah *Generalized Cross Validation* (GCV). Apabila dibandingkan dengan metode lain, misalnya *Cross Validation* (CV) dan metode *unbiased risk* (UBR)), metode GCV mempunyai sifat optimal asimtotik [8]. Titik knot optimal diperoleh dari nilai GCV yang paling minimum. Metode GCV secara umum didefinisikan sebagai berikut(Eubank, 1988).

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{(n^{-1} \text{trace}[I - A(K_1, K_2, \dots, K_r)])^2} \quad (6)$$

dengan

$$MSE(k_1, k_2, k_3, \dots, k_m) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (7)$$

D. Kriteria Model Terbaik

Selain memperhatikan GCV minimum agar didapatkan titik knot optimal, untuk memperoleh model terbaik juga memperhatikan beberapa hal yaitu nilai R^2 , jumlah parameter, dan nilai MSE. Model yang baik adalah model yang dapat menjelaskan variabilitas dari variabel respon dengan baik (R^2 tinggi) dan nilai MSE nya rendah. Selain itu, model regresi spline yang baik seharusnya memenuhi prinsip parsimoni yaitu dengan sedikit parameter yang terlibat dalam model ,tetapi sudah dapat menghasilkan R^2 yang tinggi.

E. Pengujian Parameter Model

Pengujian parameter terdiri dari dua tahapan, yaitu pengujian secara serentak kemudian dilanjutkan dengan pengujian secara parsial. Pengujian secara serentak dilakukan dengan menggunakan uji F, sementara uji parsial menggunakan uji t.

F. Pengujian Asumsi Residual

Residual dari model regresi spline harus memenuhi asumsi IIDN($0, \sigma^2$). Pemeriksaan asumsi identik dilakukan dengan uji Glejser yaitu meregresikan variabel prediktor dengan nilai mutlak residual. Apabila seluruh variabel berpengaruh prediktor tidak signifikan terhadap nilai mutlak residual, maka disimpulkan residual sudah memenuhi asumsi identik. Sementara itu untuk asumsi independen dapat diperiksa dengan membuat plot *Autocorrelation Function* (ACF). Apabila tidak terdapat lag yang keluar dari batas signifikansi maka disimpulkan residual telah memenuhi asumsi independen. Berikut diberikan rumus ACF [10].

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n-k} (e_i - \bar{e})(e_{i+k} - \bar{e})}{\sum_{i=1}^n (e_i - \bar{e})^2}, \quad k = 0, 1, 2, \dots \quad (8)$$

Asumsi lainnya yang harus terpenuhi adalah residual harus berdistribusi normal. Pemeriksaan asumsi residual dapat dilakukan dengan melakukan uji Kolmogorov-Smirnov.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Hasil Survei Sosial Ekonomi Nasional Provinsi Jawa Timur Tahun 2010, Laporan Eksekutif Keadaan Angkatan Kerja Jawa Timur 2010, Publikasi Karakteristik Penduduk Jatim menurut Hasil Survey Penduduk 2010, dan Publikasi Jawa Timur Dalam Angka 2011.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini terdiri dari satu variabel respon (y) dengan tujuh variabel yang diduga mempengaruhinya seperti yang terlampir pada Tabel 1.

C. Langkah Penelitian

Langkah-langkah analisis yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

1. Membuat *scatterplot* antara persentase kriminalitas di Jawa Timur dengan masing-masing variabel yang diduga berpengaruh untuk mengetahui bentuk pola data. Apabila terdapat komponen parametrik dan nonparametrik, maka digunakan pendekatan regresi Semiparametrik Spline.
2. Memilih titik knot optimal berdasarkan nilai GCV yang paling minimum
3. Mendapatkan model regresi Spline dengan titik knot optimal
4. Menguji signifikansi parameter regresi Spline secara serentak
5. Melakukan uji parameter regresi Spline secara parsial
6. Menguji asumsi residual
7. Menghitung nilai koefisien determinasi R^2 dan MSE
8. Menguji asumsi residual IIDN dari model regresi Spline
9. Menginterpretasikan model dan menarik kesimpulan.

IV. HASIL DAN PEMBAHASAN

A. Gambaran Umum Persentase Kriminalitas di Jawa Timur serta Variabel-Variabel yang Mempengaruhinya

Gambaran umum persentase kriminalitas di Jawa Timur serta variabel-variabel yang mempengaruhinya dijelaskan melalui statistika deskriptif. Statistika deskriptif tersebut meliputi nilai rata-rata, nilai minimum, nilai maksimum, dan standar deviasi yang teringkas dalam Tabel 2.

B. Pola Hubungan antara Persentase Kriminalitas dengan Variabel yang Diduga Mempengaruhi

Pola hubungan antara variabel respon dengan variabel-variabel yang mempengaruhinya dapat dilihat pada Gambar 1 dan Gambar 2.

Gambar 1 menunjukkan bahwa kurva regresi antara persentase kriminalitas dengan variabel x_1 dan x_2 membentuk suatu pola (parametrik) yaitu pola linier.

Gambar 2 menunjukkan bahwa kurva regresi antara persentase kriminalitas dengan kelima variabel tersebut tidak membentuk suatu pola tertentu. Hal ini menunjukkan terdapat komponen nonparametrik dalam analisis regresi spline yang akan dilakukan.

C. Model Regresi Semiparametrik Spline Linier

Berdasarkan hasil analisa *scatterplot* antara persentase kriminalitas dengan variabel-variabel yang mempengaruhinya, analisis regresi dengan pendekatan semiparametrik merupakan metode yang akan digunakan. Pendekatan semiparametrik yang digunakan dalam penelitian ini adalah regresi semiparametrik spline linier. Model yang digunakan dan dicobakan adalah regresi semiparametrik spline linier ($p=1$) dengan satu knot, dua knot, dan tiga knot ($r=1, 2, 3$).

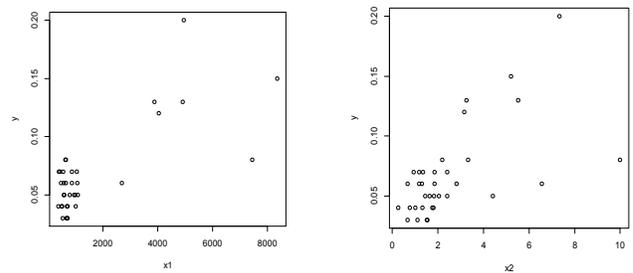
Model regresi semiparametrik Spline dengan 7 variabel yang terdiri dari dua komponen parametrik dan lima komponen nonparametrik adalah sebagai berikut.

Tabel 1.
Variabel Penelitian

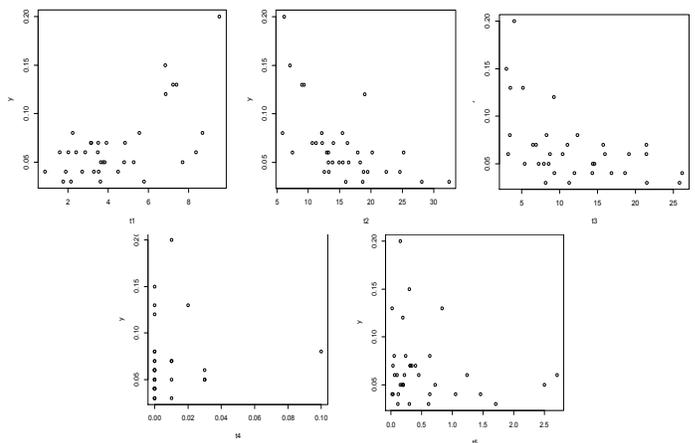
Variabel	Keterangan
y	Persentase penduduk yang melakukan tindak pidana
x_1	Kepadatan penduduk
x_2	Persentase penduduk migran
t_1	Tingkat pengangguran terbuka
t_2	Persentase penduduk miskin
t_3	Persentase penduduk yang tidak pernah sekolah
t_4	Persentase penduduk yang merupakan korban penyalahgunaan NAPZA
t_5	Persentase keluarga rentan/bermasalah

Tabel 2.
Statistika Deskriptif Persentase Kriminalitas di Jawa Timur Beserta Tujuh Faktor yang Diduga Berpengaruh

Variabel	Rata-Rata	Standar Deviasi	Minimum	Maksimum
y	0.07	0.04	0.03	0.20
x_1	1597.17	2015.75	381.00	8358.00
x_2	2.51	2.12	0.28	9.99
t_1	4.36	2.24	0.87	9.52
t_2	15.54	6.08	5.9	32.47
t_3	11.78	6.52	3.03	26.18
t_4	0.01	0.02	0.00	0.10
t_5	0.53	0.66	0.02	2.71



Gambar. 1. Scatterplot antara Persentase Kriminalitas (y) dengan Variabel x_1 dan x_2



Gambar. 2. Scatterplot antara Persentase Kriminalitas (y) dengan Variabel t_1, t_2, t_3, t_4 dan t_5

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + f(t_{i1}) + f(t_{i2}) + f(t_{i3}) + f(t_{i4}) + f(t_{i5}) + \varepsilon_i \\
 &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sum_{j=0}^p \beta_{t_{i1}}^j + \sum_{l=1}^r \beta_{p+l}(t_{i1} - k_l)_+^p + \sum_{j=0}^p \beta_{t_{i2}}^j + \\
 &\quad \sum_{l=1}^r \beta_{p+l}(t_{i2} - k_l)_+^p + \sum_{j=0}^p \beta_{t_{i3}}^j + \sum_{l=1}^r \beta_{p+l}(t_{i3} - k_l)_+^p + \sum_{j=0}^p \beta_{t_{i4}}^j + \\
 &\quad \sum_{l=1}^r \beta_{p+l}(t_{i4} - k_l)_+^p + \sum_{j=0}^p \beta_{t_{i5}}^j + \sum_{l=1}^r \beta_{p+l}(t_{i5} - k_l)_+^p + \varepsilon_i
 \end{aligned}$$

D. Pemilihan Titik Knot Optimal Dengan Satu Titik Knot

Pemilihan titik knot optimal ini dimulai dari yang paling sederhana dahulu yaitu dengan menggunakan satu titik knot. Berikut merupakan model regresi semiparametrik spline dengan satu Knot.

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 t_{i1} + \beta_4 (t_{i1} - k_1)_+ + \beta_5 t_{i2} + \beta_6 (t_{i2} - k_1)_+ + \beta_7 t_{i3} + \\
 &\quad \beta_8 (t_{i3} - k_1)_+ + \beta_9 t_{i4} + \beta_{10} (t_{i4} - k_1)_+ + \beta_{11} t_{i5} + \beta_{12} (t_{i5} - k_1)_+ + \varepsilon_i
 \end{aligned}$$

Berdasarkan hasil pada Tabel 3, GCV minimum yang dihasilkan adalah sebesar 0,00049. Knot-knot yang menghasilkan GCV minimum tersebut merupakan titik knot optimal pada pemilihan titik knot optimal dengan satu titik knot.

E. Pemilihan Titik Knot Optimal Dengan Dua Titik Knot

Model regresi semiparametrik spline dengan dua titik knot adalah sebagai berikut.

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 t_{i1} + \beta_4 (t_{i1} - k_1)_+ + \beta_5 (t_{i1} - k_2)_+ + \beta_6 t_{i2} + \\
 &\quad \beta_7 (t_{i2} - k_1)_+ + \beta_8 (t_{i2} - k_2)_+ + \beta_9 t_{i3} + \beta_{10} (t_{i3} - k_1)_+ + \beta_{11} (t_{i3} - k_2)_+ + \\
 &\quad \beta_{12} t_{i4} + \beta_{13} (t_{i4} - k_1)_+ + \beta_{14} (t_{i4} - k_2)_+ + \beta_{15} t_{i5} + \beta_{16} (t_{i5} - k_1)_+ + \beta_{17} (t_{i5} - k_2)_+ + \varepsilon_i
 \end{aligned}$$

Berdasarkan Tabel 4 didapatkan nilai GCV minimum sebesar 0,00024. Nilai tersebut lebih kecil dibandingkan GCV minimum yang dihasilkan dengan satu titik knot.

F. Pemilihan Titik Knot Optimal Dengan Tiga Titik Knot

Model regresi semiparametrik spline dengan tiga titik knot adalah sebagai berikut.

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 t_{i1} + \beta_4 (t_{i1} - k_1)_+ + \beta_5 (t_{i1} - k_2)_+ + \beta_6 (t_{i1} - k_3)_+ + \\
 &\quad \beta_7 t_{i2} + \beta_8 (t_{i2} - k_1)_+ + \beta_9 (t_{i2} - k_2)_+ + \beta_{10} (t_{i2} - k_3)_+ + \beta_{11} t_{i3} + \beta_{12} (t_{i3} - k_1)_+ + \\
 &\quad \beta_{13} (t_{i3} - k_2)_+ + \beta_{14} (t_{i3} - k_3)_+ + \beta_{15} t_{i4} + \beta_{16} (t_{i4} - k_1)_+ + \beta_{17} (t_{i4} - k_2)_+ + \\
 &\quad \beta_{18} (t_{i4} - k_3)_+ + \beta_{19} t_{i5} + \beta_{20} (t_{i5} - k_1)_+ + \beta_{21} (t_{i5} - k_2)_+ + \beta_{22} (t_{i5} - k_3)_+ + \varepsilon_i
 \end{aligned}$$

Berdasarkan hasil pada Tabel 5, GCV minimum yang didapatkan dengan menggunakan tiga titik knot adalah sebesar 0,000191. Nilai GCV ini lebih kecil dibandingkan GCV minimum yang dihasilkan dengan menggunakan dua titik knot.

G. Pemilihan Titik Knot Optimal dengan Kombinasi Knot

Pemilihan titik knot optimal juga dilakukan dengan menggunakan kombinasi knot. Hasil GCV kombinasi knot terdapat pada tabel 6.

Tabel 3.
Titik Knot Optimal dengan Satu Titik Knot

No	Knot					GCV
	t_1	t_2	t_3	t_4	t_5	
1	9.167	31.386	25.235	0.096	2.600	4.948E-04
2	8.108	28.132	22.400	0.084	2.271	7.195E-04
3	9.343	31.928	25.708	0.098	2.655	4.936E-04
4	8.814	30.301	24.290	0.092	2.490	5.374E-04
5	8.461	29.217	23.345	0.088	2.381	6.178E-04

Tabel 4.
Titik Knot Optimal dengan Dua Titik Knot

No	Knot					GCV
	t_1	t_2	t_3	t_4	t_5	
1	2.106	9.696	6.337	0.014	0.404	2.818E-04
	8.814	30.301	24.290	0.092	2.490	
2	1.929	9.153	5.865	0.012	0.349	2.769E-04
	8.990	30.843	24.763	0.094	2.545	
3	1.576	8.069	4.920	0.008	0.240	2.402E-04
	8.814	30.301	24.290	0.092	2.490	
4	1.753	8.611	5.392	0.010	0.294	2.489E-04
	8.990	30.843	24.763	0.094	2.545	
5	1.929	9.153	5.865	0.012	0.349	2.743E-04
	8.814	30.301	24.290	0.092	2.490	

Tabel 5.
Titik Knot Optimal dengan Tiga Titik Knot

No	Knot					GCV
	t_1	t_2	t_3	t_4	t_5	
	1.753	8.611	5.392	0.01	0.294	
1	7.931	27.59	21.928	0.082	2.216	2.069E-04
	9.167	31.386	25.235	0.096	2.6	
	5.989	21.625	16.731	0.059	1.612	
2	6.519	23.252	18.148	0.065	1.777	1.915E-04
	8.461	29.217	23.345	0.088	2.381	
	5.107	18.914	14.369	0.049	1.338	
3	7.049	24.879	19.566	0.071	1.941	2.171E-04
	8.108	28.132	22.4	0.084	2.271	
	7.931	27.590	21.928	0.082	2.216	
	5.813	21.083	16.259	0.057	1.557	
4	7.049	24.879	19.566	0.071	1.941	2.192E-04
	7.931	27.590	21.928	0.082	2.216	
	5.813	21.083	16.259	0.057	1.557	
5	7.049	24.879	19.566	0.071	1.941	2.222E-04
	8.990	30.843	24.763	0.094	2.545	

Berdasarkan Tabel 6 didapatkan GCV minimum yang dihasilkan yaitu sebesar 0,000188 dengan tiga titik knot pada variabel t_1, t_2, t_3, t_5 , dan satu knot pada variabel t_4 . Nilai GCV ini merupakan nilai yang paling kecil apabila dibandingkan dengan nilai GCV minimum yang dihasilkan jika menggunakan satu, dua, dan tiga titik knot. Akan tetapi terdapat kombinasi jumlah knot yang jauh lebih sederhana dengan nilai GCV yang tidak terlampaui jauh yaitu sebesar 0,000248.

H. Pemilihan model terbaik

Pemilihan model regresi semiparametrik Spline terbaik memperhatikan beberapa kriteria, diantaranya GCV minimum, R² tinggi, jumlah parameter sedikit, dan MSE rendah. Berikut merupakan resume nilai GCV, R², jumlah parameter, MSE yang dihasilkan oleh model-model dengan titik knot optimal.

Berdasarkan nilai-nilai yang terdapat pada Tabel 7, maka model yang paling memenuhi seluruh kriteria kebaikan model adalah model yang menggunakan kombinasi knot (22111). Model ini menghasilkan GCV dan MSE yang kecil, jumlah parameter sedikit dengan R² tinggi.

I. Penaksiran Parameter Model Regresi Semiparametrik Spline Linier

Berdasarkan kriteria kebaikan model, model terbaik yang dipilih adalah model regresi semiparametrik spline dengan menggunakan kombinasi jumlah knot, yaitu 2 knot pada variabel t1 dan t2, dan 1 knot pada variabel t3, t4, dan t5.

Berikut merupakan model regresi semiparametrik spline terbaik.

$$\hat{y} = -0,4432 + 0,00002x_1 + 0,00339x_2 - 0,00462x_1 + 0,00473(t_1 - 1,576)_+ + 0,22542(t_1 - 8,814)_+ + 0,05822x_2 + -0,06052(t_2 - 8,069)_+ + -0,00269(t_2 - 30,301)_+ + 0,00219x_3 + -0,020743(t_3 - 25,708)_+ + 0,35464x_4 + 0,00551(t_4 - 0,098)_+ + 0,00370x_5 + 0,00476(t_5 - 2,655)_+$$

Model tersebut menghasilkan koefisien determinasi (R²) sebesar 92,131%. Artinya, dengan hanya memuat 15 parameter model tersebut mampu menjelaskan variabilitas persentase kriminalitas di Jawa Timur sebesar 92,131% dan MSE sebesar 0,0002.

J. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter ini dilakukan secara serentak kemudian dilanjutkan dengan pengujian secara individu. Berdasarkan Tabel 8, uji serentak menghasilkan nilai statistik uji F sebesar 16,726. Sementara itu P-value yang dihasilkan sebesar 0 (kurang dari alpha 5%), sehingga diputuskan tolak H₀. Jadi kesimpulannya minimal terdapat satu variabel yang berpengaruh signifikan dalam model.

Berdasarkan Tabel 9 dapat diketahui bahwa terdapat delapan parameter atau tujuh variabel yang berpengaruh signifikan terhadap model. Dengan demikian, hanya satu variabel yang tidak berpengaruh signifikan terhadap persentase kriminalitas di Jawa Timur yaitu persentase penduduk migran (x₂).

K. Pengujian Residual

Residual model regresi spline yang telah didapatkan harus memenuhi asumsi identik, independen, dan mengikuti distribusi normal. Oleh karena itu dilakukan beberapa pengujian terhadap residual model. Pengujian pertama yaitu uji asumsi identik yang dilakukan dengan uji Glejser Tabel 10 memperlihatkan bahwa nilai F_{hitung} uji Glejser sebesar 1,1416 dengan p-value sebesar 0,384. Karena p-value lebih besar dari nilai alpha yang ditetapkan yaitu 5%, maka disimpulkan bahwa semua variabel tidak berpengaruh signifikan terhadap

Tabel 6. Titik Knot Optimal dengan Kombinasi Jumlah Knot

No	Knot					GCV
	t ₁	t ₂	t ₃	t ₄	t ₅	
1	5.989	21.625	16.731	0.098	1.612	1.880E-04
	6.519	23.252	18.148		1.777	
	8.461	29.217	23.345		2.381	
2	5.989	8.069	4.920	0.098	0.240	2.380E-04
	6.519	30.301	24.290		2.490	
3	5.989	21.625	25.708	0.098	0.240	2.420E-04
	6.519	23.252			2.490	
	8.461	29.217				
4	1.576	8.069	25.708	0.098	2.655	2.480E-04
	8.814	30.301				

Tabel 7. Nilai-Nilai Kriteria Kebaikan Model untuk Titik Knot Optimal

Kriteria kebaikan model	GCV	MSE	R ²	Jumlah parameter
1 titik knot	0,00049	0,00037	82,988	13
2 titik knot	0,00024	0,00014	94,811	18
3titik knot	0,00019	0,00013	96,690	23
kombinasi (33313)	0,00019	0,00011	96,745	21
kombinasi (22111)	0,00025	0,00019	92,131	15

Tabel 8. Tabel ANOVA Uji Serentak

Sumber Variasi	df	SS	MS	F	P-value
Regresi	14	0.04392	0.00314		
Residual	20	0.00375	0.00018	16.72606	0,00
Total	34	0.04767			

Tabel 9. Pengujian Parameter Model Spline secara Individu

Variabel	Parameter	Koefisien	t _{hitung}	P-value
Konstan	β ₀	-0.4432	-4.5114	0.000213*
	x ₁	β ₁	0.0000	8.3477
x ₂	β ₂	0.0034	0.9760	0.340705
	β ₃	-0.0046	-0.2179	0.829736
t ₁	β ₄	0.0047	0.2149	0.832015
	β ₅	0.2254	8.4518	0.000000*
	β ₆	0.0582	5.3553	0.000031*
t ₂	β ₇	-0.0605	-5.5517	0.000019*
	β ₈	-0.0027	-0.3369	0.739678
t ₃	β ₉	0.0022	2.4866	0.021849*
	β ₁₀	-0.0207	-0.6208	0.541727
t ₄	β ₁₁	0.3546	4.7724	0.000116*
	β ₁₂	0.0055	4.7518	0.000122*
t ₅	β ₁₃	0.0037	0.8976	0.380092
	β ₁₄	0.0048	3.3805	0.002972*

nilai mutlak residual. Hasil tersebut membuktikan bahwa varians residual memenuhi asumsi identik (tidak terjadi heteroskedastisitas).

Selanjutnya yaitu pengujian asumsi independen yang dilakukan dengan membuat plot *Autocorrelation Function* (ACF).

Gambar 3 menunjukkan bahwa residual memenuhi asumsi independen, terbukti dengan tidak adanya lag yang keluar dari batas toleransi.

Sementara itu, pengujian distribusi normal dilakukan dengan uji Kolmogorov-Smirnov.

Gambar 4 menunjukkan bahwa nilai uji Kolmogorov-Smirnov $D=0,0286$ lebih kecil daripada nilai $D_{tabel}=0,1497$ dengan P -value yang lebih besar daripada α 5% juga membuktikan bahwa residual model memenuhi asumsi distribusi normal.

L. Interpretasi Model Regresi Semiparametrik Spline

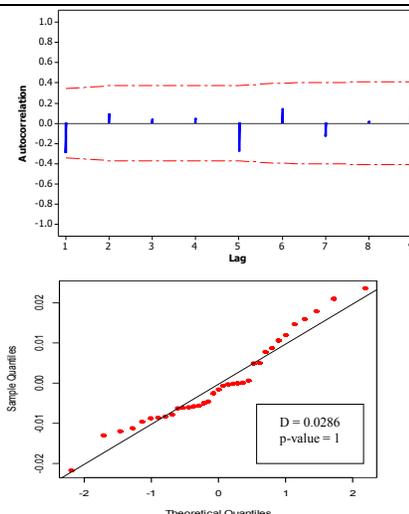
Model terbaik yang telah terpilih kemudian diinterpretasikan menurut titik knot yang signifikan. Interpretasi ini bermanfaat untuk mengetahui seberapa besar pengaruh perubahan variabel prediktor terhadap persentase kriminalitas di Jawa Timur. Dengan asumsi variabel yang lain konstan, maka interpretasi model berdasarkan masing-masing variabel adalah sebagai berikut. Apabila kepadatan penduduk naik sebesar satu satuan, maka persentase kriminalitas di Jawa Timur cenderung akan naik sebesar 0,00002%. Pada saat tingkat pengangguran terbuka lebih dari 8,814, jika tingkat pengangguran terbuka naik sebanyak satu satuan, maka persentase kriminalitas akan cenderung naik sebesar 0,22553%. Ketika persentase penduduk miskin kurang dari 8,069%, jika persentase penduduk miskin naik sebesar 1% maka persentase kriminalitas cenderung naik sebesar 0,0582%. Sementara itu pada saat persentase penduduk miskin lebih dari 8,069%, jika terjadi kenaikan persentase penduduk miskin sebesar 1% maka persentase kriminalitas cenderung turun sebesar 0,005%. Pada saat persentase penduduk yang tidak pernah sekolah kurang dari 25,078%, apabila persentase penduduk yang tidak pernah sekolah naik sebesar 1% maka persentase kriminalitas akan cenderung naik sebesar 0,00219%. ketika persentase korban penyalahgunaan NAPZA kurang dari 0,098%, jika persentase korban penyalahgunaan NAPZA bertambah 1% maka persentase kriminalitas akan cenderung naik sebesar 0,3546%. Kemudian pada saat korban penyalahgunaan NAPZA lebih dari 0,098% maka persentase kriminalitas cenderung bertambah sebanyak 0,36%. Ketika persentase keluarga bermasalah lebih dari 2,655% , apabila persentase keluarga bermasalah naik sebesar 1% maka persentase kriminalitas akan cenderung naik sebesar 0,00846%.

V. KESIMPULAN/RINGKASAN

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa setiap 10.000 penduduk di tiap kabupaten/kota di Jawa Timur, terdapat tujuh orang yang beresiko terkena tindak pidana. Kota Madiun merupakan kota dengan persentase kriminalitas tertinggi di Jawa Timur yaitu sebesar 0,2%. Terdapat enam variabel yang berpengaruh signifikan terhadap persentase kriminalitas di Jawa Timur

Tabel 10.
ANOVA Uji Glejser

Sumber Variasi	df	SS	MS	F	P-value
Regresi	14	0.00061	0.000044		
Residual	20	0.00076	0.000038	1.1416	0.384
Total	34	0.00137			



Gambar 4. *Q-Q Plot* dan Hasil Uji Kolmogorov-Smirnov Residual yaitu kepadatan penduduk, tingkat pengangguran terbuka, persentase penduduk miskin, persentase penduduk yang tidak pernah sekolah, persentase penduduk yang merupakan korban penyalahgunaan NAPZA, dan persentase keluarga bermasalah. Berdasarkan model yang telah didapatkan, keenam variabel tersebut dapat menjelaskan variabilitas persentase kriminalitas sebesar 92,131%. Saran yang dapat dilakukan untuk penelitian selanjutnya adalah agar dilakukan regresi spline dengan orde yang lebih tinggi. Selain itu Pembahasan angka kriminalitas yang lebih spesifik misal tindak kriminal jenis tertentu maupun tindak kriminal yang dilakukan oleh perempuan akan lebih menarik untuk diteliti.

DAFTAR PUSTAKA

- [1] BPS Jawa Timur., 2012, *Jawa Timur Dalam Angka 2012*, BPS Propinsi Jawa Timur, Surabaya.
- [2] Budiantara, I.N., 2006, Model Spline dengan Knots Optimal, *Jurnal Ilmu Dasar*, FMIPA Universitas Jember, 7, 77-85
- [3] Hakim, 2009, *Analisis Determinan Tingkat Kejahatan Properti di Jawa Tahun 2007.*, Fakultas Ekonomi, Universitas Indonesia, Jakarta.
- [4] Hardianto, 2009, *Analisis Faktor-Faktor yang mempengaruhi Tingkat Kriminalitas di Indonesia dari Pendekatan Ekonomi*, Fakultas Ekonomi, Universitas Katolik Parahyangan, Bandung.
- [5] Putra. P. P., 2008, *Identifikasi Faktor-Faktor yang Mempengaruhi Tindak Kriminal Berdasarkan Karakteristik Pelaku Kriminal dengan Menggunakan Metode Pohon Klasifikasi*, Jurusan Matematika, Universitas Andalas, Padang.
- [6] Dwimayanti. I., 2006, *Aplikasi Regresi Spline Untuk Memperkirakan Tingkat Fertilitas Berdasarkan Rata-Rata Paritas*, Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya.
- [7] Eubank, R., 1988, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.
- [8] Wahba, G. 1990. *Spline Models For Observational Data*. Pennsylvania: SIAM.
- [9] Oktaviana, 2011, *Regresi Spline Birespon untuk Memodelkan Kadar Gula Darah Penderita Diabetes Melitus*, Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya
- [10] Wei, W.W.S. (2006). *Time Series Analysis: Univariate and Multivariate* (2nd ed.). USA: Pearson Education, Inc.