

Pemodelan Kejahatan di Indonesia dengan Metode Regresi Nonparametrik Spline Truncated

Atma Aurelia Caroline Cindy dan Madu Ratna
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: madu_r@statistika.its.ac.id

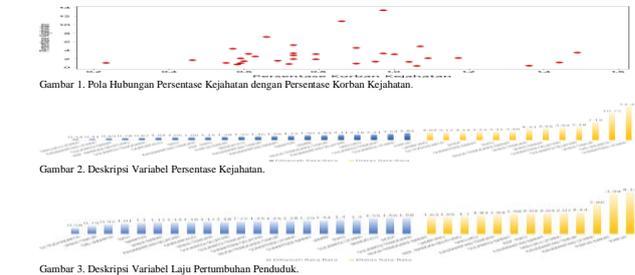
Abstrak—*The Global Initiative Against Transnational Organized Crime (The Global Initiative)* pada tahun 2021 mencatat indeks kejahatan terhadap 193 negara yang bergabung sebagai anggota PBB, yang dimana Indonesia tercatat berada di tingkat ke-25 dengan skor 6,38. Sedangkan berdasarkan publikasi BPS tahun 2021, jumlah kejahatan di Indonesia dari tahun 2018-2020 mengalami penurunan. Untuk mencapai tujuan pembangunan berkelanjutan berdasarkan program SDGs, perlu diketahui jumlah kejahatan pada suatu wilayah, sebab dengan meningkatnya kejahatan berarti semakin rendah tingkat keamanan pada suatu wilayah. Penelitian ini akan memodelkan kejahatan menggunakan regresi nonparametrik spline truncated sebab hubungan antara variabel respon yaitu kejahatan dengan kelima variabel prediktor yang diduga mempengaruhi tidak berpola serta terdapat perubahan pola hubungan pada sub interval tertentu. Berdasarkan penelitian, model regresi nonparametrik spline terbaik adalah dengan menggunakan kombinasi titik knot (1,2,1,3,1) dengan menghasilkan koefisien determinasi sebesar 88,28% dan diperoleh tiga variabel yang berpengaruh signifikan terhadap kejahatan di Indonesia tahun 2020 yaitu variabel laju pertumbuhan penduduk, IPM, dan rata-rata lama sekolah.

Kata Kunci—Indonesia, Kejahatan, Knot, Regresi Non Parametrik Spline.

I. PENDAHULUAN

SETIAP negara menghadapi permasalahan berupa kejahatan, yang dimana besar kecilnya hambatan dari kejahatan berpengaruh terhadap pambangunan dari satu negara. Proses pembangunan akan berjalan lancar jika kejahatan dapat ditekan serendah-rendahnya, yang dimana hal tersebut juga dipengaruhi oleh peran aktif dan dukungan dari masyarakat. Banyak kerugian yang disebabkan adanya tindak kejahatan, baik secara ekonomi, fisik, moral, bahkan psikologis. Berdasarkan sudut pandang ekonomi, efisiensi biaya kejahatan, maka kejahatan lebih baik dilakukan pencegahan daripada membiarkannya terjadi kemudian diberantas. Peningkatan kejahatan memaksa suatu negara untuk meningkatkan pengeluaran untuk keamanan dan ketertiban publik, sehingga menyebabkan penggunaan sumber daya yang tidak efisien, baik di sektor publik maupun swasta. Sehingga untuk menekan persentase kejahatan, perlu dilakukan suatu studi atau penelitian mengenai kejahatan dan faktor-faktor yang mempengaruhinya.

Kejahatan atau kejahatanitas merupakan suatu bentuk tindakan yang melanggar peraturan perundang-undangan yang telah ditetapkan oleh pemerintah dan menyimpang dari norma-norma sosial serta meresahkan masyarakat [1]. Istilah kejahatan atau kejahatan juga bahwasanya memiliki pengertian secara yuridis (formal) dan sosiologis. Secara yuridis, kejahatan merupakan bentuk perilaku yang bertentangan dengan moral kemanusiaan, merugikan



masyarakat, melanggar hukum serta undang-undang. Secara sosiologis, kejahatan merupakan semua bentuk ucapan, perbuatan, dan tingkah laku yang secara ekonomis, politis, dan sosial-psikologis dapat sangat merugikan masyarakat [2].

Beberapa penelitian mengenai persentase kejahatan pernah dilakukan, yaitu dengan menggunakan metode Regresi Semiparametrik Spline di Jawa Timur dengan faktor-faktor yang memengaruhi persentase kejahatan adalah kepadatan penduduk, tingkat pengangguran terbuka, persentase penduduk miskin, penduduk yang menggunakan NAPZA, persentase keluarga bermasalah, dan persentase penduduk yang tidak pernah sekolah. Berikutnya, penelitian dilakukan pemodelan Regresi Data Panel pada data jumlah kejahatan di Indonesia dengan faktor-faktor yang memengaruhi adalah persentase penduduk miskin, IPM, rata-rata lama sekolah.

Penelitian mengenai jumlah kejahatan di Indonesia ini akan dimodelkan dengan menggunakan regresi nonparametrik *spline truncated*. Metode tersebut digunakan berdasarkan melihat plot pada variabel-variabel yang diduga mempengaruhi jumlah kejahatan di Indonesia, sebab plot data yang terbentuk tidak membentuk suatu pola tertentu atau dapat dikatakan menyebar, sehingga pola datanya tidak diketahui. Model ini digunakan karena *spline* memiliki kelebihan, yaitu *spline* merupakan potongan polinomial yang memiliki sifat tersegmen. Keunggulan metode regresi nonparametrik *spline truncated* sehingga digunakan pada penelitian ini karena memiliki fleksibilitas yang tinggi dan mampu memodelkan data yang memiliki jika data berubah-ubah pada interval tertentu dan model cenderung mencari sendiri estimasinya kemanapun data tersebut bergerak. Keunggulan tersebut terjadi karena dalam *spline* terdapat titik-titik knot, yang dimana titik knot merupakan titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data [3]. Hal tersebut dapat dilihat dari pola hubungan antara variabel jumlah kejahatan dengan salah satu variabel prediktor yaitu persentase korban kejahatan yang terdapat pada Gambar 1.

Penelitian ini menggunakan beberapa variabel yang diduga mempengaruhi jumlah kejahatan di Indonesia yaitu persentase laju pertumbuhan penduduk penduduk, indeks pembangunan manusia, persentase penduduk miskin, rata-

Tabel 1.

Struktur Data Penelitian							
No	Provinsi	Y	X ₁	X ₂	X ₃	X ₄	X ₅
1	Provinsi Aceh	Y ₁	X _{1,1}	X _{2,1}	X _{3,1}	X _{4,1}	X _{5,1}
2	Provinsi Sumatera Utara	Y ₂	X _{1,2}	X _{2,2}	X _{3,2}	X _{4,2}	X _{5,2}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	Provinsi Papua	Y ₃₄	X _{1,34}	X _{2,34}	X _{3,34}	X _{4,34}	X _{5,34}

Tabel 2.

Variabel Penelitian				
No	Variabel	Keterangan	Satuan	
1	Y	Kejahatan	Rasio	
2	X ₁	Laju Pertumbuhan Penduduk	Rasio	
3	X ₂	Indeks Pembangunan Manusia	Rasio	
4	X ₃	Persentase Penduduk Miskin	Rasio	
5	X ₄	Rata-Rata Lama Sekolah	Rasio	
6	X ₅	Persentase Korban Kejahatan	Rasio	

Tabel 3.

Statistika Deskriptif dari Setiap Variabel Penelitian				
Variabel	Rata-Rata	Varians	Minimum	Maksimum
Y	2,94	7,79	0,34	13,34
X ₁	1,62	0,57	0,58	4,13
X ₂	71,08	15,23	60,44	80,77
X ₃	10,81	29,30	4,45	26,80
X ₄	9,08	0,69	6,96	11,17
X ₅	0,83	0,08	0,23	1,49

Tabel 4.

Nilai Titik Knot dan GCV Pada Satu Knot						
No	X ₁	X ₂	X ₃	X ₄	X ₅	GCV
1	2,445	71,121	16,193	9,172	0,892	6,179
2	2,505	71,466	16,572	9,243	0,913	6,196
3	2,566	71,811	16,951	9,315	0,935	6,122
4	2,626	72,156	17,308	9,386	0,9561	6,279
5	2,686	72,500	17,708	9,457	0,9774	6,617

Tabel 5.

Nilai Titik Knot dan GCV Pada Dua Knot						
No	X ₁	X ₂	X ₃	X ₄	X ₅	GCV
1	2,391	70,813	15,853	9,108	0,873	6,207
	4,13	80,77	26,8	11,17	1,49	
2	2,464	71,227	16,309	9,194	0,899	6,207
3	2,536	71,642	16,765	9,280	0,924	5,455
4	2,609	71,227	16,309	9,194	0,899	6,108
5	2,686	72,057	17,221	9,366	0,95	6,108
	2,464	71,227	16,309	9,194	0,899	6,672
	2,681	72,472	17,678	9,452	0,976	
	2,464	71,227	16,309	9,194	0,899	
	2,753	72,887	18,134	9,538	1,001	7,259

Tabel 6.

Nilai Titik Knot dan GCV Pada Tiga Knot						
No	X ₁	X ₂	X ₃	X ₄	X ₅	GCV
1	2,898	73,717	19,046	9,709	1,053	4,656
	2,971	74,132	19,502	9,795	1,079	
	3,985	79,940	25,888	10,998	1,437	
2	2,898	73,717	19,046	9,709	1,053	4,703
	2,971	74,132	19,502	9,795	1,079	
	4,056	80,355	26,344	11,084	1,464	
3	2,898	73,717	19,046	9,709	1,053	4,615
	3,043	74,547	19,502	9,881	1,104	
	3,116	74,961	20,414	9,967	1,13	
4	2,898	73,717	19,046	9,709	1,053	3,528
	3,043	74,547	19,958	9,881	1,104	
	3,188	75,376	20,870	10,053	1,156	
5	2,898	73,717	19,046	9,709	1,053	4,110
	3,043	74,547	19,958	9,881	1,104	
	3,261	75,791	21,327	10,139	1,181	

Tabel 7.

Nilai Titik Knot dan GCV Pada Kombinasi Knot						
No	X ₁	X ₂	X ₃	X ₄	X ₅	GCV
1	2,566	71,227	16,951	9,315	1,053	6,886
		71,642			1,156	
2	2,566	71,227	16,951	9,194	0,935	6,345
		71,642		9,280	0,899	
3	2,566	71,227	16,951	9,194	1,053	6,266
		71,642		9,280	1,104	
4	2,566	71,227	16,951	9,194	1,156	7,504
		71,642		9,280	1,053	
5	2,566	71,227	16,951	9,709	0,935	2,565
		71,642		9,881	1,104	
				10,053		

rata lama sekolah dan persentase korban kejahatan. Se jauh ini, masih belum ada penelitian terkait kejahatan di Indonesia menggunakan regresi nonparametrik *spline truncated*. Hasil dari penelitian ini diharapkan dapat memberikan pemodelan yang baik dan mengetahui faktor-faktor yang mempengaruhi kejahatan di Indonesia, sehingga dapat dijadikan salah satu pertimbangan bagi pemerintah untuk menekan kejahatan dari setiap provinsi yang terdapat di Indonesia.

II. URAIAN PENELITIAN

A. Statistika Deskriptif

Statistika deskriptif adalah suatu metode yang berkaitan dengan pengumpulan data kemudian digunakan untuk menyajikan gugus data sehingga diperoleh data yang informatif dan menarik. Namun perlu diketahui juga bahwa statistika deskriptif hanya memberikan informasi mengenai data yang dipunyai tetapi tidak dapat menarik kesimpulan dari masalah yang sedang ditangani [4]. Mean adalah rasio dari total pengamatan dengan banyak pengamatan atau titik keseimbangan massa dari gugus pengamatan.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

keterangan :

\bar{x} : mean

x_i : pengamatan ke- i , $i=1,2,\dots,n$



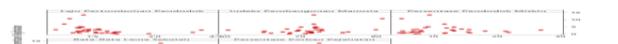
Gambar 2. Deskripsi Variabel PersentasePenduduk Miskin.



Gambar 5. Deskripsi Variabel Indeks Pembangunan Manusia.



Gambar 3. Deskripsi Variabel Persentase Korban Kejahatan.



Gambar 4. Scatterplot Antara Persentase Kejahatan dengan Variabel yang Diduga Mempengaruhi.



Gambar 5. Plot ACF dari Residual.



Gambar 6. Hasil Normal Probability Plot Residual pada Uji Kolmogorov Smirnov.



Gambar 7. Persebaran Provinsi Berdasarkan Laju Pertumbuhan Penduduk.



Gambar 8. Persebaran Provinsi Berdasarkan Indeks Pembangunan Manusia.



Gambar 9. Persebaran Provinsi Berdasarkan Rata-Rata Lama Sekolah.

n : banyak pengamatan

Varians adalah rata-rata berbobot dari kuadrat jarak setiap nilai data terhadap pusat data dengan rumus berikut :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \tag{2}$$

keterangan :

s^2 : varians

\bar{x} : mean

x_i : pengamatan ke- i , $i=1,2,\dots,n$

n : banyak pengamatan

B. Analisis Regresi

Analisis regresi dapat digunakan untuk memprediksi suatu nilai dari variabel dependen berdasarkan nilai-nilai yang terdapat pada variabel independen. Sebelum dilakukan analisis regresi, perlu dilihat terlebih dahulu pola data berdasarkan diagram pencar. Jika terdapat tren linier, kuadrat, kubik, polinomial, eksponensial dan sigmoid maka data tersebut dianalisis dengan pendekatan regresi parametrik. Dan jika tren atau pola hubungan antara variabel respon dan prediktor membentuk pola sembarang maka digunakan pendekatan regresi nonparametrik [5].

C. Regresi Nonparametrik

Regresi nonparametrik merupakan regresi dengan tidak adanya asumsi mengenai bentuk kurva regresi, atau tidak terdapat informasi masa lalu lengkap tentang bentuk pola data, sehingga dapat berbentuk apa saja, baik linier maupun nonlinier. Sebab itu kurva regresinya hanya diasumsikan *smooth* (mulus) yang berarti termuat dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik sangat mempertahankan fleksibilitasnya [6]. Bentuk umum persamaan dari model regresi nonparametrik yaitu:

$$y_i = f(x_i) + \varepsilon_i \tag{3}$$

keterangan :

y_i : variabel respon ke- i

x_i : variabel prediktor

$f(x_i)$: fungsi regresi yang tidak mengikuti pola tertentu

ε_i : error random ke- i yang diasumsikan berdistribusi $N(0,\sigma^2)$

i : 1,2, ...,n, dengan n adalah banyak pengamatan

D. Regresi Nonparametrik Spline Truncated

Spline pada regresi nonparametrik memiliki sifat fleksibel dan mempunyai kemampuan untuk mengestimasi perilaku data yang cenderung berbeda pada interval berlainan. Spline

merupakan potongan (*truncated*) polinomial tersegmen yang kontinu sehingga memiliki kemampuan menyesuaikan diri lebih efektif dari karakteristik data. Pendekatan spline memiliki keunggulan dalam mengestimasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan relatif mulus. Titik knot merupakan titik perpaduan yang menunjukkan perubahan perilaku kurva pada selang yang berbeda [7].

E. Estimasi Parameter

Estimasi adalah suatu proses yang berguna untuk menghasilkan suatu nilai tertentu terhadap suatu parameter. *Ordinary Least Square* (OLS) merupakan salah satu metode yang biasa digunakan pada pengestimasian parameter model regresi nonparametrik *spline truncated*. Metode ini mengestimasi parameter dengan cara meminimumkan jumlah kuadrat residual. Estimasi parameter $\hat{\beta}$ diberikan pada persamaan sebagai berikut:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{4}$$

F. Pemilihan Titik Knot Optimal

Model regresi spline terbaik merupakan model yang memiliki titik knot optimal. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola kurva pada interval yang berlainan. Titik knot ialah titik perpaduan yang memperlihatkan perubahan perilaku data atau fungsi pada interval tertentu. Metode yang baik untuk memilih titik knot optimal yaitu GCV. Metode GCV mempunyai sifat optimal asimtotik jika dibandingkan dengan metode lain, misalnya *Cross Validation* (CV) dan *Unbiased Risk* (UBR) [8]. Metode CGV dapat dituliskan pada persamaan sebagai berikut:

$$GCV(k) = \frac{MSE(k)}{[n^{-1}trace(I-A)]^2} \tag{5}$$

Dengan ukuran kinerja atas penduga yang sederhana adalah kuadrat dari sisaan yang dirata-rata.

$$MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \tag{6}$$

keterangan :

- I : matriks identitas
- n : banyak pengamatan
- k : (k_1, k_2, \dots, k_r) titik-titik knot
- $\hat{f}(x_i)$: penduga fungsi $f(x)$

G. Pengujian Parameter Model

Pengujian parameter model dilakukan untuk mengetahui signifikansi pengaruh model dan kelayakan model pada penggambaran variabel prediktor terhadap variabel respon.

1) Uji Secara Serentak Parameter Model Regresi

Uji serentak adalah uji signifikansi parameter model secara keseluruhan atau untuk mengetahui apakah semua variabel prediktor yang dimasukkan ke dalam model memberikan pengaruh secara bersama-sama dengan hipotesis uji serentak sebagai berikut.

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

$$H_1 : \text{minimal terdapat satu } \beta_q \neq 0, q = 1, 2, \dots, p + k$$

Dengan Statistik uji

$$F_{hitung} = \frac{MSR}{MSE} \tag{7}$$

Daerah penolakan H_0 jika $F_{hitung} > F_{\alpha(p+k; n-(p+k)-1)}$ atau $p\text{-value} < \alpha$. Jika H_0 ditolak maka disimpulkan bahwa minimal terdapat satu parameter pada model regresi yang signifikan terhadap model.

2) Pengujian Parsial Parameter Model Regresi

Uji parsial atau individu yaitu uji untuk mengetahui pengaruh dari masing-masing variabel prediktor secara

individu terhadap variabel respon [9]. Berikut merupakan hipotesis untuk uji parsial pada model regresi.

Perumusan Hipotesis:

$$H_0 : \beta_j = 0, j = 1, 2, \dots, p + k$$

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

Dengan statistik uji

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, j = 1, 2, \dots, p + k \tag{8}$$

Daerah penolakan adalah H_0 ditolak jika nilai $|t_{hitung}| > t_{(\frac{\alpha}{2}, n-(p+k)-1)}$ atau $p\text{-value} < \alpha$. Kesimpulan yang diperoleh jika tolak H_0 adalah parameter berpengaruh signifikan terhadap variabel respon.

H. Koefisien Determinasi

Semakin tinggi nilai yang dihasilkan koefisien determinasi maka semakin baik juga variabel-variabel prediktor pada suatu model yang dimana menjelaskan variabilitas dari variabel respon. Dengan menggunakan koefisien determinasi (R^2) maka dapat menentukan kebaikan dari model regresi yang telah dibuat. Persamaan yang digunakan untuk mendapatkan nilai koefisien determinasi yaitu

$$R^2 = \frac{SS_{regresi}}{SS_{total}} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{9}$$

I. Asumsi Residual

Untuk mengetahui residual yang dihasilkan pada model regresi nonparametrik telah memenuhi asumsi IIDN (Identik, Independen, Distribusi Normal) maka dapat melakukan uji asumsi residual (*goodness of fit*).

1) Asumsi Identik

Dengan dilakukannya uji asumsi identik dapat mengetahui terdapat homogenitas pada variansi residual model regresi. Jika variansi antar residual tidak homogen maka terjadi kasus heteroskedastisitas dan dilakukan pendeteksian dengan uji *Glejser* [10]. Uji *Glejser* dilakukan dengan cara meregresikan nilai mutlak residual dengan variabel prediktornya. Hipotesis yang digunakan yaitu.

$$H_0 : \text{var}(\epsilon_1) = \text{var}(\epsilon_2) = \dots = \text{var}(\epsilon_n) = \sigma^2 \text{ (residual identik)}$$

$$H_1 : \text{minimal ada satu } \text{var}(\epsilon_i) \neq \sigma^2; i = 1, 2, \dots, n \text{ (residual tidak identik)}$$

Statistik uji yang digunakan adalah sebagai berikut:

$$F_{hitung} = \frac{\sum_{i=1}^n ((|\epsilon_i| - |\bar{\epsilon}|)^2) / (v-1)}{\sum_{i=1}^n (|\epsilon_i| - |\bar{\epsilon}|)^2 / (n-v)} \tag{10}$$

H_0 ditolak jika $F_{hitung} < F_{\alpha; (v-1, n-v)}$ atau $p\text{-value} > \alpha$, dan v adalah banyak parameter model. Dengan adanya tolak H_0 maka dapat disimpulkan bahwa terdapat indikasi adanya kasus heteroskedastisitas yang berarti asumsi identik tidak terpenuhi.

2) Asumsi Independen

Pemeriksaan independen dilakukan untuk mengetahui autokorelasi atau tidak. Dengan tidak terdapat autokorelasi antar residual pada nilai kovarian antara ϵ_i dan ϵ_{i-s} sama dengan nol maka dapat memenuhi uji asumsi independen. Salah satu cara mendeteksi autokorelasi pada residual adalah dengan plot ACF. Nilai koefisien ACF pada lag ke- s digunakan rumus persamaan sebagai berikut [11]:

$$\hat{\rho}_s = \frac{\sum_{i=s+1}^n (\epsilon_{i-s} - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2} \tag{11}$$

Keterangan:

- $\hat{\rho}_s$: korelasi antara ϵ_i dan ϵ_{i-s}
- s : lag ke- s (1, 2, ...)

Interval konfidensi sebesar $(1-\alpha) \times 100\%$ digunakan untuk mendeteksi asumsi independen dengan persamaan sebagai berikut:

$$-t_{\frac{\alpha}{2}, n-1} SE(\hat{\rho}_s) < \rho_s < t_{\frac{\alpha}{2}, n-1} SE(\hat{\rho}_s) \tag{12}$$

dengan nilai $SE(\hat{\rho}_s)$ seperti pada persamaan berikut:

Tabel 8.

Perbandingan Nilai GCV Minimum Antar Model Knot	
Model	GCV Minimum
Satu titik knot	6.122079
Dua titik knot	5.455276
Tiga titik knot	3.528053
Kombinasi titik knot (1,2,1,1)	2.565188

Tabel 9.

Hasil Pengujian Secara Serentak					
Sumber	Derajat Bebas	SS	MS	F-hat	P-Value
Regresi	13	226.890	17.453	11.585	0.0000013
Error	20	30.130	1.506		
Total	33	25.02			

Tabel 10.

Hasil Pengujian Secara Parsial					
Variabel	Parameter	Estimasi	t-hitung	P-value	Keputusan
Konstan	β0	9.106	0.497	0.624	Tidak Signifikan
	β1	-1.88	-2.341	0.029	Signifikan
	β2	1.887	0.972	0.342	Tidak Signifikan
X1	β3	0.089	0.321	0.75	Tidak Signifikan
	β4	6.115	3.165	0.004	Signifikan
	β5	-7.183	-3.913	0.008	Signifikan
X2	β6	-0.014	-0.181	0.858	Tidak Signifikan
	β7	0.134	0.37	0.715	Tidak Signifikan
	β8	-1.379	-2.442	0.024	Signifikan
X3	β9	75.361	7.025	0.0000008	Signifikan
	β10	-153.922	-6.792	0.000001	Signifikan
	β11	94.794	6.529	0.000002	Signifikan
X4	β12	1.596	0.925	0.366	Tidak Signifikan
	β13	1.494	0.465	0.647	Tidak Signifikan

Tabel 11.

Hasil Pengujian Glejser				
Sumber	DF	SS	F-hat	P-value
Regresi	13	2.84088	0.2185293	0.5460131
Error	20	8.00454	0.4002272	
Total	33	10.8454		

$$SE(\hat{\rho}_s) = \sqrt{\frac{1+2\sum_{i=1}^{s-1}(\rho_s)^2}{n}} \tag{13}$$

Apabila terdapat minimal satu nilai ACF ($\hat{\rho}_s$) yang keluar dari interval konfidensi, maka diindikasikan terdapat autokorelasi antar residual. Sebaliknya, jika semua nilai ACF berada di dalam batas interval konfidensi maka tidak terdapat autokorelasi atau bisa dikatakan asumsi independen terpenuhi.

3) Asumsi Normalitas

Uji asumsi normalitas atau distribusi normal dilakukan untuk mengetahui pada model regresi sudah berdistribusi normal atau tidak. Pengujian ini dapat dilakukan dengan *normal probability plot* dan juga *Kolomogorof Smirnof* dengan hipotesis sebagai berikut [12]:

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

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

$$D = \sup|S(x) - F_0(x)| \tag{14}$$

Daerah penolakan yaitu $|D| > q_{(1-\alpha,n)}$ atau $P\text{-value} < \alpha$, dimana nilai $q_{(1-\alpha,n)}$ ditentukan dari tabel *Kolomogorof Smirnof*. Keputusan yang diambil dari tolak H_0 adalah residual tidak berdistribusi normal.

J. Kejahatan

Secara yuridis formal, kejahatan merupakan bentuk dari tingkah laku yang bertentangan dengan moral kemanusiaan (*immorill*), merugikan masyarakat dan melanggar hukum serta undang-undang pidana. Sedangkan secara sosiologis, kejahatan merupakan bentuk ucapan, perbuatan, dan tingkah laku yang secara ekonomis, politis, dan sosial-psikologis sangat merugikan masyarakat, melanggar norma-norma Susila, dan menyerang keselamatan warga masyarakat.

Berdasarkan Publikasi Statistik Kejahatan 2021 yang dilansirkan oleh BPS (2021) menjelaskan bahwa tindak kejahatan/kejahatan atau pelanggaran merupakan perbuatan seseorang yang dapat diancam hukuman berdasarkan KUHP dan Undang-Undang serta peraturan lainnya yang berlaku di Indonesia. Setiap peristiwa yang dilaporkan masyarakat kepada Polri, atau peristiwa dimana pelakunya tertangan tangan oleh kepolisian merupakan peristiwa yang dilaporkan mengenai kejahatan atau kejahatan. Terdapat beberapa unsur tindak pidana yaitu [13]:

1. Kelakuan dan akibat atau perbuatan
2. Hal ihwal atau keadaan yang memberikan pidana
3. Keadaan tambahan yang memberatkan pidana
4. Unsur melawan hukum yang objektif

5. Unsur melawan hukum yang subjektif

III. METODOLOGI PENELITIAN

A. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari BPS Indonesia dengan judul Statistik Kejahatan 2021. Adapun data yang digunakan berdasarkan data pada tahun 2020. Variabel respon penelitian ini adalah kejahatan, 34 observasi yaitu banyaknya provinsi yang ada di Indonesia, serta 5 variabel prediktor merupakan variabel yang diduga mempengaruhi persentase jumlah kejahatan di Indonesia.

B. Struktur Penelitian

Struktur data penelitian yang digunakan digunakan dalam penelitian ini diuraikan dalam Tabel 1.

C. Variabel Penelitian

Variabel penelitian yang digunakan berdasarkan pada penjelasan yang ada di sumber data terkait 1 variabel respon berupa kejahatan dan 5 variabel prediktor yang diduga berpengaruh terhadap jumlah kejahatan di Indonesia pada tahun 2020 tertera pada Tabel 2.

D. Langkah Analisis

Untuk menjawab permasalahan pada penelitian ini, dilakukan langkah analisis sebagai berikut.

1. Merumuskan masalah dan melakukan studi literatur terkait jumlah kejahatan di Indonesia tahun 2020.
2. Mengumpulkan data yang akan dianalisis.
3. Melakukan analisis deskriptif untuk mendeskripsikan karakteristik jumlah kejahatan berdasarkan faktor-faktor yang diduga mempengaruhinya.
4. Mengidentifikasi pola hubungan yang terbentuk antara variabel respon terhadap masing-masing variabel prediktor dengan menggunakan *scatter plot*.
5. Melakukan pemodelan jumlah kejahatan berdasarkan faktor-faktor yang diduga berpengaruh pada setiap provinsi di Indonesia dengan menggunakan regresi nonparametrik *spline truncated*.
6. Memodelkan variabel respon menggunakan pendekatan *spline titik knot* 1, 2, 3, dan kombinasi.
7. Hasil nilai GCV yang paling minimum digunakan sebagai titik knot optimal.
8. Membuat model regresi nonparametrik *spline truncated* yang terbaik menggunakan titik knot optimal.
9. Melakukan pengujian signifikan parameter secara serentak dan parsial
10. Melakukan pengujian asumsi residual identik, independen dan berdistribusi normal (IIDN) dari model regresi nonparametrik *spline truncated* yang terbentuk.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteris Data

Analisis deskriptif pada data kejahatan dan faktor-faktor yang diduga mempengaruhinya di Indonesia pada tahun 2020 bertujuan untuk melihat karakteristik kejahatan dari setiap provinsi di Indonesia berdasarkan masing-masing variabel prediktor. Hasil karakteristik data kejahatan terdapat pada Tabel 3.

Berdasarkan Tabel 3, diketahui bahwa variabel respon (Y) yang merupakan persentase kejahatan memiliki rata-rata

sebesar 2,94 dengan nilai varians sebesar 7,79. Nilai varians yang besar menunjukkan bahwa kejahatan di Indonesia tahun 2020 tersebar secara tidak merata. Masih banyaknya provinsi di Indonesia yang memiliki kejahatan yang tinggi akan mengakibatkan tidak tercapainya atau terhambatnya program SDGs pada poin ke-16 mengenai kedamaian serta rendahnya tingkat keamanan pada suatu provinsi tersebut. Berikut ini merupakan bentuk visual melalui grafik batang dari persentase kejahatan di Indonesia pada tahun 2020.

Gambar 2 menunjukkan bahwa provinsi yang memiliki persentase kejahatan di Indonesia tahun 2020 yang paling rendah adalah Provinsi Maluku Utara sebesar 0,34, sedangkan provinsi dengan persentase kejahatan tertinggi terjadi di Provinsi Sumatera Utara dengan nilai 13,34. Persentase kejahatan menunjukkan angka yang beragam, hal ini terlihat dari besarnya perbedaan antara provinsi yang memiliki angka tertinggi dan terendah. Perbedaan tersebut dapat terjadi karena beberapa factor seperti laju pertumbuhan penduduk, indeks pembangunan manusia, persentase penduduk miskin dan lainnya yang tidak sama atau tidak merata.

1) *Karakteristik Faktor Laju Pertumbuhan Penduduk di Indonesia Tahun 2020*

Laju Pertumbuhan Penduduk atau variabel X_1 merupakan salah satu faktor yang diduga mempengaruhi persentase kejahatan di Indonesia tahun 2020. Nilai rata-rata yang diperoleh Laju Pertumbuhan Penduduk di Indonesia pada tahun 2020 yaitu sebesar 1,62 persen dengan varians 0,57. Nilai minimum dan maksimum dari variabel laju pertumbuhan penduduk masing-masing sebesar 0,58 persen yang terdapat di Provinsi DI Yogyakarta dan 4,13 persen di Provinsi Papua. Gambar 3 menunjukkan deskripsi variabel laju pertumbuhan penduduk.

Salah satu faktor yang mempengaruhi tinggi rendahnya nilai laju pertumbuhan penduduk adalah kelahiran, kematian dan migrasi. Gambar 4 menunjukkan deskripsi variabel penduduk miskin.

2) *Karakteristik Faktor Indeks Pembangunan Manusia di Indonesia Tahun 2020*

Indeks Pembangunan Manusia atau variabel X_2 merupakan salah satu faktor yang diduga mempengaruhi persentase kejahatan di Indonesia tahun 2020. Tabel 3 memberikan informasi bahwa nilai rata-rata yang diperoleh indeks pembangunan manusia di Indonesia pada tahun 2020 yaitu sebesar 71,08 persen dengan varians 15,23.

Gambar 5 menjelaskan bahwa Provinsi Papua merupakan provinsi yang memiliki nilai indeks pembangunan manusia paling kecil dengan nilai 60,44 persen, sedangkan Provinsi DKI Jakarta merupakan provinsi dengan nilai indeks pembangunan manusia paling besar dengan nilai 80,77. Tinggi rendahnya Nilai IPM pada suatu provinsi dipengaruhi oleh 3 komponen, yaitu tingkat pendidikan, tingkat kesehatan, dan standar hidup yang layak.

3) *Karakteristik Faktor Persentase Penduduk Miskin di Indonesia Tahun 2020*

Persentase penduduk miskin atau variabel X_3 merupakan salah satu faktor yang diduga mempengaruhi persentase kejahatan di Indonesia tahun 2020. Tabel 3 memberikan informasi bahwa nilai rata-rata yang diperoleh persentase

penduduk miskin di Indonesia pada tahun 2020 yaitu sebesar 10,81 persen dengan varians 29,30.

Pada Gambar 4 menjelaskan bahwa Provinsi Bali merupakan provinsi minimum persentase penduduk miskin dengan nilai 4,45 persen dan provinsi dengan persentase penduduk miskin tertinggi adalah Provinsi Papua dengan nilai 26,8 persen. Dari data BPS persentase penduduk miskin merupakan persentase yang berada di bawah Garis Kemiskinan (GK). Garis kemiskinan mencerminkan nilai rupiah pengeluaran minimum yang diperlukan seseorang untuk memenuhi kebutuhan pokok hidupnya selama sebulan, baik kebutuhan makanan maupun non-makanan.

4) *Karakteristik Faktor Rata-Rata Lama Sekolah di Indonesia Tahun 2020*

Rata-rata lama sekolah atau variabel X_4 merupakan salah satu faktor yang diduga mempengaruhi persentase kejahatan di Indonesia tahun 2020. Tabel 3 memberikan informasi bahwa nilai rata-rata yang diperoleh rata-rata lama sekolah di Indonesia pada tahun 2020 yaitu sebesar 9,08 persen dengan varians 0,69.

Dari Gambar 6 menjelaskan bahwa Provinsi Papua merupakan provinsi dengan nilai rata-rata lama sekolah paling rendah dengan nilai 6,96 persen dan provinsi dengan rata-rata lama sekolah tertinggi adalah Provinsi DKI Jakarta dengan nilai 11,17 persen. Hal ini menunjukkan bahwa masih terdapat provinsi di Indonesia yang belum bisa menyelesaikan target pendidikan minimal sekolah, sebab pembangunan sarana pendidikan berorientasi di provinsi dengan sarana pendidikan yang memadai, sedangkan terbatas untuk provinsi yang memiliki wilayah dengan sarana pendidikan yang kurang memadai.

5) *Karakteristik Faktor Persentase Korban Kejahatan di Indonesia Tahun 2020*

Persentase korban kejahatan atau variabel X_5 merupakan salah satu faktor yang diduga mempengaruhi persentase kejahatan di Indonesia tahun 2020. Tabel 3 memberikan informasi bahwa nilai rata-rata yang diperoleh persentase korban kejahatan di Indonesia pada tahun 2020 yaitu sebesar 0,83 persen dengan varians 0,08.

Gambar 6 menjelaskan bahwa Provinsi Bali merupakan provinsi dengan persentase korban kejahatan yang paling rendah dengan nilai 0,23 persen, sedangkan provinsi dengan persentase korban kejahatan yang paling tinggi terdapat pada Provinsi Nusa Tenggara Barat dengan nilai 1,49 persen. Semakin tinggi persentase korban kejahatan pada suatu provinsi menunjukkan bahwa semakin banyak tindak kejahatan yang mengakibatkan adanya korban kejahatan.

B. *Pola Hubungan Kejahatan Terhadap Faktor-Faktor yang Diduga Mempengaruhi*

Hasil dari diagram *scatterplot* dapat menunjukkan pola hubungan dari variabel respon terhadap setiap variabel prediktor apakah terdapat pola hubungan tertentu atau tidak terdapat pola hubungan tertentu secara visual. Berikut merupakan pola hubungan antara Persentase Kejahatan dengan kelima variabel yang diduga mempengaruhinya.

Berdasarkan Gambar 7 dapat diketahui bahwa pola hubungan antara Persentase Kejahatan di Indonesia tahun 2020 dengan semua variabel yang diduga mempengaruhinya pola yang terbentuk yaitu tidak mengikuti atau tidak

membentuk suatu pola tertentu. Hal ini dapat diketahui dengan adanya sebaran plot yang menyebar secara acak, sehingga menunjukkan bahwa semua variabel prediktor merupakan komponen nonparametrik dan dapat dilakukan pemodelan dengan menggunakan metode regresi nonparametrik *spline truncated*.

C. Pemodelan IPM Menggunakan Regresi Nonparametrik Spline

1) Regresi Nonparametrik Spline truncated Satu Titik Knot

Pemodelan Regresi Nonparametrik *Spline truncated* dengan satu titik knot pada variabel yang diduga mempengaruhi Persentase Kejahatan di Indonesia yaitu sebagai berikut.

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2(x_1 - k_1)_+ + \beta_3x_2 + \beta_4(x_2 - k_2)_+ + \beta_5x_3 + \beta_6(x_3 - k_3)_+ + \beta_7x_4 + \beta_8(x_4 - k_4)_+ + \beta_9x_5 + \beta_{10}(x_5 - k_5)_+$$

Hasil perhitungan nilai knot pada setiap variabel prediktor dan GCV minimum untuk model regresi nonparametrik *spline truncated* menggunakan satu knot yang disajikan pada Tabel 4.

Berdasarkan Tabel 4 menunjukkan nilai titik knot dan GCV yang paling minimum pada satu knot yaitu 6,122 dengan titik optimal pada masing-masing variabel prediktor yaitu

- X₁ : K₁ = 2,565593
- X₂ : K₂ = 71,81102
- X₃ : K₃ = 16,95085
- X₄ : K₄ = 9,314746
- X₅ : K₅ = 0,934746

2) Regresi Nonparametrik Spline truncated Dua Titik Knot

Estimasi dari model regresi nonparametrik *spline truncated* terhadap persentase kejahatan di Indonesia menggunakan dua titik knot sebagai berikut

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2(x_1 - k_1)_+ + \beta_3(x_1 - k_2)_+ + \beta_4x_2 + \beta_5(x_2 - k_3)_+ + \beta_6(x_2 - k_4)_+ + \beta_7x_3 + \beta_8(x_3 - k_5)_+ + \beta_9(x_3 - k_6)_+ + \beta_{10}x_4 + \beta_{11}(x_4 - k_7)_+ + \beta_{12}(x_4 - k_8)_+ + \beta_{13}x_5 + \beta_{14}(x_5 - k_9)_+ + \beta_{15}(x_5 - k_{10})_+$$

Berdasarkan persamaan model regresi nonparametrik *spline truncated* dengan dua titik knot, diperoleh hasil perhitungan nilai GCV dengan dua titik knot dari setiap variabel prediktor yang diduga mempengaruhi persentase kejahatan di Indonesia dan disajikan pada Tabel 5.

Berdasarkan Tabel 5 menunjukkan nilai titik knot dan GCV yang paling minimum pada dua knot yaitu 5,455 dengan titik optimal dari setiap variabel prediktor yaitu

- X₁ : (K₁ = 2,463673 K₂ = 2,536122)
- X₂ : (K₃ = 71,22735 K₄ = 71,64224)
- X₃ : (K₅ = 16,30918 K₆ = 16,76531)
- X₄ : (K₇ = 9,193878 K₈ = 9,279796)
- X₅ : (K₉ = 0,898571 K₁₀ = 0,898571)

3) Regresi Nonparametrik Spline truncated Tiga Titik Knot

Setelah dilakukan pemilihan nilai GCV minimum dari satu titik knot dan dua titik knot, maka dilanjutkan dengan pemilihan titik knot optimum dengan menggunakan tiga titik knot dengan persamaan sebagai berikut.

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2(x_1 - k_1)_+ + \beta_3(x_1 - k_2)_+ + \beta_4(x_1 - k_3)_+ + \beta_5x_2 + \beta_6(x_2 - k_4)_+ + \beta_7(x_2 - k_5)_+ + \beta_8(x_2 - k_6)_+ + \beta_9x_3 + \beta_{10}(x_3 - k_7)_+ + \beta_{11}(x_3 - k_8)_+ + \beta_{12}(x_3 - k_9)_+ + \beta_{13}x_4 + \beta_{14}(x_4 - k_{10})_+ + \beta_{15}(x_4 - k_{11})_+ + \beta_{16}(x_4 - k_{12})_+ + \beta_{17}x_5 + \beta_{18}(x_5 - k_{13})_+ + \beta_{19}(x_5 - k_{14})_+ + \beta_{20}(x_5 - k_{15})_+$$

Berdasarkan hasil pengolahan, Tabel 6 merupakan nilai titik knot dan GCV pada model regresi nonparametrik *spline truncated* dengan tiga titik knot.

Tabel 6 memberikan informasi bahwa nilai titik knot dan GCV yang paling minimum pada tiga knot yaitu 3,528 dengan titik optimal dari masing-masing variabel prediktor yaitu

- X₁ : (K₁ = 2,898367 K₂ = 3,043265 K₃ = 3,188163)
- X₂ : (K₄ = 73,71673 K₅ = 74,54653 K₆ = 75,37633)

- X₃ : (K₇ = 19,04592 K₈ = 19,95816 K₉ = 20,87041)
- X₄ : (K₁₀ = 9,709388 K₁₁ = 9,881224 K₁₂ = 10,05306)
- X₅ : (K₁₃ = 1,052857 K₁₄ = 1,104286 K₁₅ = 1,155714)

4) Regresi Nonparametrik Spline truncated Kombinasi Titik Knot

Tabel 7 merupakan nilai GCV dan knot yang dihasilkan menggunakan kombinasi titik knot pada model regresi nonparametrik *spline truncated*.

Berdasarkan Tabel 7 nilai GCV yang paling minimum adalah 2,565 yaitu kombinasi (1,2,1,3,1) dengan lokasi titik knot optimum untuk masing-masing variabel adalah sebagai berikut.

- X₁ : (K₁ = 2,565593)
- X₂ : (K₂ = 71,22735 K₃ = 71,64224)
- X₃ : (K₄ = 16,95085)
- X₄ : (K₅ = 9,709388 K₆ = 9,881224 K₇ = 10,05306)
- X₅ : (K₈ = 0,9347458)

Estimasi model regresi nonparametrik *spline truncated* pada persentase kejahatan di Indonesia tahun 2020 dengan menggunakan kombinasi knot yang terbentuk yaitu.

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2(x_1 - k_1)_+ + \beta_3x_2 + \beta_4(x_2 - k_2)_+ + \beta_5(x_2 - k_3)_+ + \beta_6x_3 + \beta_7(x_3 - k_4)_+ + \beta_8x_4 + \beta_9(x_4 - k_5)_+ + \beta_{10}(x_4 - k_6)_+ + \beta_{11}(x_4 - k_7)_+ + \beta_{12}x_5 + \beta_{13}(x_5 - k_8)_+$$

D. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan melakukan perbandingan nilai GCV, Tabel 8 merupakan hasil perbandingan antar model berdasarkan hasil analisis sebelumnya.

Berdasarkan Tabel 8 yang merupakan perbandingan dari setiap model titik knot yang digunakan yaitu satu, dua, tiga, maupun kombinasi titik knot, diketahui bahwa model titik knot optimal atau model yang memiliki nilai GCV paling minimum terdapat pada model kombinasi titik knot. Adapun model kombinasi titik knot tersebut adalah dengan kombinasi 1,2,1,3,1 dan nilai GCV yang dihasilkan yaitu sebesar 2,56519.

E. Pengujian Signifikansi Parameter Model Pengujian

1) Uji Serentak

Pengujian parameter secara serentak dilakukan untuk mengetahui signifikansi parameter model secara keseluruhan dengan melibatkan seluruh variabel prediktor. Adapun hipotesis yang digunakan adalah sebagai berikut.

- H₀ : β₁ = β₂ = ... = β₁₃ = 0
- H₁ : minimal terdapat satu β_j ≠ 0, j=1,2,...,13

Hasil pengujian secara serentak menggunakan ANOVA disajikan dalam Tabel 9. Berdasarkan hasil ANOVA terdapat pada Tabel 9 dengan menggunakan taraf signifikansi (α) sebesar 0,05 nilai dari F_(0,05;13,20) adalah sebesar 1.312, sehingga keputusannya adalah tolak H₀ sebab nilai Fhitung = 11,585 > F_(0,05;13,20). Hal ini menunjukkan bahwa terdapat paling tidak satu variabel yang memberikan pengaruh signifikan terhadap persentase kejahatan di Indonesia pada tahun 2020. Selanjutnya dilakukan pengujian secara parsial untuk melihat pengaruh dari setiap variabel prediktor terhadap variabel respon secara individu.

2) Uji Parsial

Pengujian secara parsial atau individu dilakukan apabila pada pengujian parameter didapatkan keputusan tolak H₀, yang dimana model secara serentak terdapat minimal satu parameter yang signifikan. Pengujian secara parsial menggunakan hipotesis sebagai berikut.

- H₀ : β_j = 0
- H₁ : β_j ≠ 0, j=1,2,...,13

Hasil pada pengujian parameter secara parsial ditunjukkan pada Tabel 10. Berdasarkan Tabel 10 hasil pengujian parameter secara parsial pada model regresi nonparametrik *spline truncated* yang terbentuk, menunjukkan terdapat beberapa parameter pada variabel prediktor yang tidak signifikan terhadap angka kejahatan di Indonesia. Hal ini disebabkan beberapa parameter tersebut memiliki nilai $|t_{hitung}|$ lebih kecil dari nilai $t_{0,025;(13)}$ (2,16) atau dengan menggunakan taraf signifikan (α) sebesar 0,05, variabel dikatakan signifikan apabila terdapat salah satu p-value pada parameter memiliki nilai kurang dari 0,05 maka keputusan tolak H_0 , sehingga dapat diartikan bahwa kelima parameter tersebut tidak signifikan terhadap model yang diperoleh. Apabila terdapat minimal satu parameter yang signifikan dari satu variabel prediktor, maka variabel tersebut dikatakan berpengaruh signifikan terhadap respon meskipun parameter lain tidak signifikan.

F. Asumsi Residual

1) Asumsi Identik

Asumsi identik terpenuhi apabila tidak terdapat indikasi terjadinya heterokedasitas yang dapat mengakibatkan kerugian pada efisiensi estimator, dengan kata lain varians harus homogen. Pemeriksaan asumsi identik dilakukan dengan menggunakan uji Glejser. Uji Glejser dilakukan dengan meregresikan antara harga mutlak dari hasil tabel ANOVA untuk uji Glejser dengan hipotesis sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{34}^2 = \sigma^2$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2; i = 1, 2, \dots, 34$$

Berdasarkan hasil tabel ANOVA yang diperoleh dari pengujian Glejser pada Tabel 11 diketahui bahwa F_{hitung} sebesar $0,55 < F_{(0,05;13;20)}$ (2,25), sehingga diperoleh keputusan gagal tolak H_0 . Hal ini diperkuat dengan nilai p-value $> 0,05$. Sehingga dapat diartikan bahwa residual model regresi nonparametrik *spline truncated* tidak terjadi heterokedasitas dan menunjukkan bahwa variansi residual homogen serta telah memenuhi asumsi identik.

2) Asumsi Residual Independen

Asumsi independent dari model regresi mengharuskan tidak ada korelasi antar residual. Salah satu cara mendeteksi adanya autokorelasi adalah dengan menggunakan ACF. Adapun hipotesis yang digunakan pada asumsi independent yaitu sebagai berikut.

$$H_0 : \rho = 0 \text{ (Tidak terjadi autokorelasi)}$$

$$H_1 : \rho \neq 0 \text{ (Terjadi autokorelasi)}$$

Gambar 8 merupakan hasil pendeteksi autokorelasi dengan plot ACF. Berdasarkan Gambar 8 terlihat bahwa tidak terdapat nilai korelasi yang dihasilkan keluar dari batas signifikansi pada semua *lag* sehingga disimpulkan bahwa tidak ada nilai autokorelasi yang keluar dari batas signifikansi. Hal tersebut menunjukkan asumsi independent pada residual telah terpenuhi.

3) Asumsi Residual Berdistribusi Normal

Pemeriksaan asumsi normalitas dilakukan untuk mengetahui apakah residual berdistribusi normal atau tidak. Pada pemeriksaan asumsi normalitas dilakukan dengan menggunakan Uji *Kolmogorov-Smirnov*, uji ini merupakan salah satu metode yang dilakukan untuk mengetahui suatu data telah mengikuti distribusi tertentu dengan hipotesis yang digunakan sebagai berikut.

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

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

Gambar 9 merupakan hasil pengujian asumsi residual normal menggunakan uji *Kolmogorov-Smirnov* yang ditampilkan pada *Normal Probability Plot*.

Berdasarkan Gambar 9 memberikan informasi bahwa $|D|(0,075) < D_\alpha(0,227)$ atau nilai P-value dari hasil pengujian yaitu $> 0,150$ yang dimana nilai tersebut lebih besar dari α sebesar 0,05. Sehingga berdasarkan hasil tersebut, keputusan yang diambil adalah gagal tolak H_0 yang berarti residual mengikuti distribusi normal atau asumsi residual berdistribusi normal terpenuhi. Hasil Pemeriksaan asumsi residual didapati bahwa residual telah memenuhi asumsi IIDN.

G. Koefisien Determinasi

Nilai koefisien determinasi (R^2) menunjukkan seberapa besar kebaikan model regresi dalam menjelaskan variabilitas kejahatan di Indonesia tahun 2020. Berdasarkan perhitungan yang telah dilakukan, nilai koefisien determinasi yang dihasilkan oleh model regresi nonparametrik *spline truncated* adalah sebesar 88.28 persen.

H. Interpretasi Tiap Variabel Signifikan Pada Model Regresi Nonparametrik Spline Truncated

Interpretasi model dilakukan dengan tujuan untuk mengetahui seberapa besarnya pengaruh masing-masing variabel terhadap kejahatan di Indonesia pada tahun 2020. Berikut merupakan interpretasi model dari setiap variabel yang signifikan terhadap kejahatan di Indonesia.

Pengaruh hubungan variabel laju pertumbuhan penduduk (X_1) terhadap persentase kejahatan di Indonesia pada tahun 2020 dengan mengasumsikan variabel lainnya dianggap konstan dapat dituliskan sebagai berikut

$$\hat{y} = -1.880x_1 + 1.887(x_1 - 2.566)^{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} -1.880x_1 & x_1 < 2.566 \\ 0.007x_1 - 4,842 & x_1 \geq 2.566 \end{cases}$$

Berdasarkan model tersebut, diketahui bahwa untuk provinsi-provinsi yang memiliki laju pertumbuhan penduduk (X_1) kurang dari 2,566, berarti apabila laju pertumbuhan penduduk di provinsi-provinsi ini bertambah satu persen maka persentase jumlah kejahatan di Indonesia akan mengalami penurunan sebesar 1,880. Adapun provinsi-provinsi yang mengalami hal ini terdiri dari 31 provinsi.

Sedangkan pada provinsi-provinsi yang memiliki laju pertumbuhan penduduk (X_1) lebih dari atau sama dengan 2,566, berarti apabila laju pertumbuhan penduduk di provinsi-provinsi ini bertambah satu persen maka persentase jumlah kejahatan di Indonesia akan meningkat sebesar 0,007. Adapun provinsi-provinsi yang mengalami hal ini terdiri dari 3 provinsi. Pengelompokkan provinsi berdasarkan variabel laju pertumbuhan penduduk yang mempengaruhi persentase kejahatan di Indonesia tahun 2020 secara visual disajikan menggunakan Gambar 10.

Pengaruh hubungan variabel indeks pembangunan manusia (X_2) terhadap persentase kejahatan di Indonesia pada tahun 2020 dengan mengasumsikan variabel lainnya dianggap konstan dapat dituliskan sebagai berikut

$$\hat{y} = 0.089x_2 + 6.115(x_2 - 71.227)^{\frac{1}{4}} - 7.183(x_2 - 71.642)^{\frac{1}{4}}$$

$$\hat{y} = \begin{cases} 0.089x_2 & ; x_2 < 71.227 \\ 6,204x_2 - 435,553 ; 71.227 \leq x_2 < 71.642 \\ -0,979x_2 + 79,051 ; & x_2 \geq 71.642 \end{cases}$$

Dari model yang terbentuk, dapat diinterpretasikan bahwa provinsi-provinsi yang memiliki indeks pembangunan manusia (X_2) kurang dari 71,227 persen, berarti apabila indeks pembangunan manusia di provinsi-provinsi tersebut

mengalami peningkatan satu persen, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan meningkat sebesar 0,089 persen. Provinsi yang mengalami hal tersebut terdiri dari 15 provinsi.

Ketika provinsi-provinsi di Indonesia yang memiliki indeks pembangunan manusia (X_2) berada diantara 71,227 hingga 71,642 persen, berarti apabila indeks pembangunan manusia di provinsi-provinsi tersebut mengalami peningkatan satu persen, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan meningkat sebesar 6,204 persen. Adapun provinsi-provinsi yang mengalami hal tersebut terdiri dari 4 provinsi.

Kemudian pada provinsi-provinsi di Indonesia yang memiliki indeks pembangunan manusia (X_2) berada lebih atau sama dengan 71,642 persen, berarti apabila indeks pembangunan manusia di provinsi-provinsi tersebut mengalami peningkatan satu persen, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan mengalami penurunan sebesar 0,979 provins. Provinsi yang mengalami hal ini terdiri dari 15 Provinsi. Pengelompokan provinsi berdasarkan variabel indeks pembangunan manusia yang mempengaruhi persentase kejahatan di Indonesia tahun 2020 secara visual disajikan menggunakan Gambar 11.

Pengaruh hubungan variabel rata-rata lama sekolah (X_4) terhadap persentase kejahatan di Indonesia pada tahun 2020 dengan mengasumsikan variabel lainnya dianggap konstan dapat dituliskan sebagai berikut

$$\hat{y} = -1.379x_4 + 75.361(x_4 - 9.709)_+^1 - 153.922(x_4 - 9.881)_+^1 + 94.794(x_4 - 10.053)_+^1$$

$$\hat{y} = \begin{cases} -1.379x_4 & ; x_4 < 9.709 \\ 73.982x_4 - 731.680 & ; 9.709 \leq x_4 < 9.881 \\ -79.94x_4 + 789.223 & ; 9.881 \leq x_4 < 10.053 \\ 14.854x_4 - 163.741 & ; x_4 \geq 10.053 \end{cases}$$

Berdasarkan model diatas, untuk provinsi-provinsi yang memiliki rata-rata lama sekolah (X_4) kurang dari 9,709, berarti apabila rata-rata lama sekolah di provinsi-provinsi tersebut dinaikkan satu persen dengan variabel lain konstan, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan berkurang sebesar 1,379 persen, dimana provinsi yang mengalami hal ini terdiri dari 25 provinsi.

Kemudian, ketika provinsi-provinsi di Indonesia memiliki rata-rata lama sekolah (X_4) berada dikisaran 9,709 hingga 9,881, berarti apabila rata-rata lama sekolah di provinsi-provinsi tersebut dinaikkan satu persen dengan variabel lain konstan, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan bertambah sebesar 73,982 persen. Terdapat dua provinsi yang mengalami hal tersebut yaitu Provinsi Aceh dan Sulawesi Utara.

Pada ketika provinsi-provinsi di Indonesia memiliki rata-rata lama sekolah (X_4) berada dikisaran 9,881 hingga 10,053 serta, berarti apabila rata-rata lama sekolah di provinsi-provinsi tersebut dinaikkan satu persen dengan variabel lain konstan, maka persentase jumlah kejahatan di provinsi-provinsi tersebut akan berkurang sebesar 79,94 persen. Terdapat 4 provinsi yang mengalami hal tersebut.

Ketika provinsi-provinsi di Indonesia memiliki rata-rata lama sekolah (X_4) lebih atau sama dengan 10,053 persen, apabila rata-rata lama sekolah di provinsi-provinsi tersebut dinaikkan satu persen dengan variabel lain konstan, maka persentase jumlah kejahatan di provinsi-provinsi tersebut

akan bertambah sebesar 14,854 persen. Terdapat tiga provinsi yang mengalami hal tersebut. Pengelompokan provinsi berdasarkan variabel rata-rata lama sekolah di Indonesia tahun 2020 secara visual disajikan menggunakan Gambar 12.

V. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dijelaskan pada bab sebelumnya, didapatkan beberapa kesimpulan sebagai berikut: pada tahun 2020 rata-rata persentase kejahatan di Indonesia adalah sebesar 2,94 dengan varians sebesar 7,79. Persentase kejahatan terendah pada tahun 2020 terdapat pada Provinsi Maluku Utara dengan nilai 0,34, sedangkan nilai tertinggi tertinggi terjadi di Provinsi Sumatera Utara dengan nilai 13,34.

Model regresi nonparametrik *spline truncated* yang digunakan untuk memodelkan persentase kejahatan di Indonesia tahun 2020 adalah dengan kombinasi knot 1,2,1,3,1 dengan nilai GCV sebesar 2,56188. Residual model telah memenuhi asumsi IIDN. Selain itu nilai koefisien determinasi yang diperoleh adalah sebesar 88,28%, sedangkan sisanya dipengaruhi oleh variabel lain yang tidak diteliti. Berikut merupakan model regresi nonparametrik *spline truncated* yang diperoleh.:

$$\hat{y} = 9.107 - 1.880x_1 + 1.887(x_1 - 2.566)_+^1 + 0.089x_2 - 71.227)_+^1 + 6.115(x_2 - 7.183(x_2 - 71.642)_+^1 - 0.014x_3 + 0.134(x_3 - 16.951)_+^1 - 1.379x_4 + 75.361(x_4 - 9.709)_+^1 - 153.922(x_4 - 9.881)_+^1 + 94.794(x_4 - 10.053)_+^1 + 1.596x_5 + 1.494(x_5 - 0.935)_+^1$$

Berdasarkan hasil pengujian parameter secara parsial menunjukkan terdapat beberapa parameter yang tidak signifikan terhadap persentase kejahatan di Indonesia tahun 2020 yaitu variabel persentase penduduk miskin dan korban kejahatan. Sedangkan variabel laju pertumbuhan penduduk, indeks pembangunan manusia, dan rata-rata lama sekolah memberikan pengaruh yang signifikan terhadap persentase kejahatan di Indonesia pada tahun 2020.

DAFTAR PUSTAKA

- [1] K. Kartono, *Patologi Sosial*. Jakarta: Rajawali, 1992.
- [2] F. N. Hardianto, "Analisis faktor-faktor yang mempengaruhi tingkat kriminalitas di Indonesia dari pendekatan ekonomi," *Bina Ekon.*, vol. 13, no. 2, 2009.
- [3] I. N. Budiantara, "Metode U, GLM, GV, dan GCV dalam regresi nonparametrik spline.," *Maj. Ilm. Himpun. Mat. Indones.*, vol. 6, no. 285--290, 2000.
- [4] R. E. Walpole, *Pengantar Statistika*, 3rd ed. Jakarta: Gramedia Pustaka Utama, 1995.
- [5] I. N. Budiantara, "Spline dalam Regresi Nonparametrik: Sebuah Pemodelan Statistik Masa Kini dan Masa Mendatang," Departemen Statistika: Institut Teknologi Sepuluh Nopember, 2009.
- [6] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 3rd ed. New York: Marcel Dekker, 1999.
- [7] W. Härdle, *Applied Nonparametric Regression*. London: Cambridge University Press, 1990.
- [8] G. Wahba, *Spline Models for Observational Data*. Pennsylvania: SIAM, 1990.
- [9] N. R. Draper and H. Smith, *Applied Regression Analysis*. New Jersey: John Wiley & Sons, 1998.
- [10] D. N. Gujarati, *Basic Econometric Analysis*. New Jersey: McGraw-Hill, 2004.
- [11] W. Wei, *Time Series Analysis Univariate and Multivariate*. London: Pearson education, 2006.
- [12] W. N. Daniel, *Applied Nonparametric Statistics*, 2nd ed. Boston: Houghton Mifflin, 1999.
- [13] S. Moeljatno, *Asas-asas Hukum Pidana*. Jakarta: Rineka Cipta, 2002.