

Regresi Nonparametrik Multivariabel dengan Pendekatan *Spline Truncated* Pada Kasus *Tuberculosis*

MUHAMMAD HUNAIPi PRATAMA¹, SIFRIYANI², DARNAH ANDI NOHE³

^{1,2,3}Program Studi Statistika, Jurusan Matematik, Fakultas MIPA, Universitas Mulawarman, Samarinda, Indonesia

e-mail: sifriyani@fmipa.unmul.ac.id

ABSTRAK

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan respon ketika tidak diketahui bentuk fungsi regresinya dimana kurva regresi hanya diasumsikan *smooth*. *Spline truncated* adalah suatu pendekatan ke arah pencocokan data dengan tetap memperhitungkan kemulusan kurva yang memiliki sifat tersegmentasi kontinu sehingga memberikan fleksibilitas dan efektif menyesuaikan diri dalam menjelaskan karakteristik lokal dari fungsi data. Tujuan penelitian ini adalah untuk mendapat model regresi nonparameterik multivariabel dengan pendekatan *spline truncated* dan mengetahui faktor-faktor yang berpengaruh terhadap jumlah kasus *tuberculosis* di Kalimantan Timur, Kalimantan Selatan dan Kalimantan Barat tahun 2019 menggunakan data yang didapatkan dari Badan Pusat Statistik. Hasil menunjukkan bahwa model regresi nonparametrik *spline truncated* terbaik dalam memodelkan jumlah kasus *tuberculosis* adalah dengan tiga titik knot. Model ini memiliki nilai *generalized cross validation* (GCV) yang paling minimum sebesar 40.813,24, dengan nilai R^2 sebesar 95,84%. Berdasarkan hasil pengujian signifikansi parameter model, disimpulkan bahwa faktor-faktor yang memengaruhi jumlah kasus *tuberculosis* di Kalimantan Timur, Kalimantan Selatan dan Kalimantan Barat tahun 2019 adalah persentase penduduk miskin, persentase tenaga kesehatan terlatih, persentase keluarga memperoleh sanitasi, dan persentase gizi buruk masyarakat.

Kata Kunci: Regresi Nonparametrik, *Spline Truncated*, GCV, *Tuberculosis*.

ABSTRACT

Nonparametric regression is a statistical method used to determine the pattern of the relationship between the predictor variable and the response when the form of the regression function is unknown where the regression curve is only assumed to be smooth. The truncated spline is an approach towards data matching while taking into account the smoothness of the curve which has continuous segmented properties so as to provide flexibility and effectively adapt to explain the local characteristics of the data function. The purpose of this study was to obtain a multivariable nonparametric regression model with a spline truncated approach and find out the factors that influence the number of tuberculosis cases in East Kalimantan, South Kalimantan and West Kalimantan in 2019 using the data from Indonesian Central Bureau of Statistics. The results show that the spline truncated nonparametric regression model is the best in modeling the number of cases. tuberculosis is with three point knots. This model has the minimum generalized cross validation (GCV) value of 40,813.24, with an R^2 value of 95.84%. Based on the results of testing the significance of the model parameters, it was concluded that the factors that influenced the number of tuberculosis cases in East Kalimantan, South Kalimantan and West Kalimantan in 2019 were the percentage of poor people, the percentage of trained health workers, the percentage of families receiving sanitation, and the percentage of poor nutrition in the community.

keywords: Nonparametric regression, *Spline Truncated*, GCV, *Tuberculosis*.

1. PENDAHULUAN

Analisis regresi merupakan metode statistika untuk memodelkan hubungan keterkaitan antara dua variabel atau lebih. Suatu fungsi pada model regresi dapat didekati dengan pendekatan parametrik, pendekatan semiparametrik, dan pendekatan nonparametrik.

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan respon ketika tidak diperoleh informasi sebelumnya tentang bentuk fungsi regresinya atau tidak diketahui bentuk kurva regresinya (Putra, et. al. 2015). Secara umum fungsi regresi dapat dinyatakan seperti pada persamaan berikut ini (Eubank, 1999):

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

Pendekatan *spline* adalah suatu pendekatan kearah pencocokan data dengan tetap memperhitungkan kemulusan kurva. *Spline* merupakan potongan (*piecewise*) polinomial yang memiliki sifat tersegmen kontinu sehingga efektif menjelaskan karakteristik lokal dari fungsi data (Eubank, 1988). Secara umum bentuk fungsi *spline truncated* berorde m dengan titik knot K_1, K_2, \dots, K_q dapat dinyatakan dengan persamaan berikut:

$$f(x_i) = \sum_{j=1}^m \beta_j x_i^j + \sum_{h=1}^q \beta_{m+h} (x_i - K_h)_+^m \quad (2)$$

dengan

$$(x_i - K_h)_+^m = \begin{cases} (x_i - K_h)^m, & x_i \geq K_h \\ 0, & x_i < K_h \end{cases} \quad (3)$$

Keterangan:

- β_j = Koefisien parameter regresi untuk fungsi polinomial ke j , $j = 1, 2, \dots, m$ dengan m adalah banyaknya orde.
- x_i = Variabel prediktor untuk observasi ke- i , $i = 1, \dots, n$, n adalah banyak amatan.
- β_{m+h} = Koefisien parameter regresi untuk fungsi truncated orde m dan h adalah titik knot, dimana $h = 1, 2, \dots, q$
- K_h = Knot ke- h , dengan $h = 1, 2, \dots, q$

nilai m pada persamaan (3) merupakan derajat polinomial. Kurva polinomial derajat satu disebut dengan kurva linier, derajat dua disebut dengan kurva kuadratik, derajat tiga disebut dengan kurva kubik, dan seterusnya.

Orde pada model ditentukan berdasarkan pola umum yang terjadi pada data, pendekatan *spline* mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik knot. Titik knot merupakan titik yang terdapat pada perubahan pola perilaku fungsi (Eubank, 1999), dalam penentuan titik knot optimal terdapat beberapa metode yang dilakukan salah satunya metode *Generalized Cross Validation* (GCV). Penggunaan metode GCV dalam pemilihan titik knot merupakan yang terbaik dikarenakan metode tersebut memiliki sifat optimal asimtotik (Wahba, 1990). Berikut ini merupakan persamaan untuk nilai GCV sebagai berikut (Eubank, 1988):

$$GCV(K_q) = \frac{MSE(K_q)}{(n^{-1}tr[\mathbf{I} - \mathbf{A}(K_q)])^2} \quad (4)$$

Keterangan:

- K_q = Knot ke- q
- MSE = Mean Standard Error
- tr = trace, jumlah bilangan pada diagonal utama matriks
- $\mathbf{A}(K_q)$ = Matriks hat, dirumuskan dengan $[\mathbf{X}[\mathbf{K}]^T \mathbf{X}[\mathbf{K}]]^{-1} \mathbf{X}[\mathbf{K}]^T$

Tuberculosis adalah suatu penyakit infeksi menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, yang dapat menyerang berbagai organ terutama paru-paru. Penyakit ini bila tidak diobati atau pengobatannya tidak tuntas dapat menimbulkan

komplikasi berbahaya hingga menyebabkan kematian (Kementerian Kesehatan, RI, 2015). Penyakit *tuberculosis* masih menjadi momok bagi masyarakat Indonesia, tercatat hingga bulan Juli 2020 oleh kementerian kesehatan Indonesia jumlah kasus *tuberculosis* di Indonesia mencapai 845 ribu namun yang baru ditemukan sekitar 69 persen, bahkan sebelum pandemi *covid-19* terjadi, secara global Indonesia berada di posisi ketiga setelah India dan China untuk kasus *tuberculosis* terbanyak di dunia. Penyebab dari kenaikan kasus *tuberculosis* yang signifikan disebabkan karena perilaku masyarakat yang masih belum bisa menjaga kebersihan ditambah banyaknya titik-titik perumahan yang masih masuk dalam kawasan kumuh.

2. METODE PENELITIAN

Data diperoleh dari lembaga atau instansi melalui *website* resmi Badan Pusat Statistik dan Dinas Kesehatan. Periode data yang digunakan dalam penelitian ini adalah tahun 2019. Unit analisis yang digunakan dalam penelitian ini adalah kabupaten/kota di provinsi Kalimantan Barat, Kalimantan Selatan, dan Kalimantan Timur, sehingga total ada 37 observasi yang digunakan dalam penelitian ini. Adapun variabel yang digunakan dalam penelitian ini dapat dilihat pada tabel 1.

Tabel 1 Variabel Penelitian

Variabel	Nama Variabel	Skala
y	Jumlah Kasus <i>Tuberculosis</i>	Rasio
x_1	Persentase Penduduk Miskin	Rasio
x_2	Persentase Tenaga Kesehatan Terlatih	Rasio
x_3	Persentase Keluarga Memperoleh Sanitasi yang Layak	Rasio
x_4	Persentase Gizi Buruk Masyarakat	Rasio

Metode yang digunakan pada penelitian ini adalah regresi nonparametrik *multivariable* dengan pendekatan *spline truncated*. Metode penelitian ini terdiri dari beberapa tahapan analisis, berikut tahapan analisis pada penelitian ini:

1. Mengumpulkan data sekunder mengenai kasus penyakit *tuberculosis* di Kalimantan Timur, Kalimantan Selatan dan Kalimantan Barat tahun 2019 beserta faktor-faktor yang memengaruhinya.
2. Melakukan analisis statistika deskriptif pada variabel respon dan variabel prediktor yang bertujuan dapat mengetahui karakteristik mengenai kasus penyakit *tuberculosis*
3. Mengidentifikasi pola data yang terbentuk dari kasus penyakit *tuberculosis* dengan setiap faktor yang memengaruhinya menggunakan *scatterplot*.
4. Melakukan pendeteksian multikolinieritas pada variabel prediktor.
5. Memodelkan data mengenai kasus penyakit *tuberculosis* beserta semua faktor-faktor yang memengaruhinya menggunakan regresi nonparametrik *spline multivariable* dengan satu, dua, tiga, dan kombinasi titik knot.
6. Memilih titik knot dengan menggunakan *Generalized Cross Validation* (GCV)
7. Memodelkan kasus penyakit *tuberculosis* dari model yang terpilih dengan menggunakan regresi nonparametrik *spline* pada titik knot optimal yang terpilih.
8. Melakukan uji signifikansi parameter secara simultan dan dilanjutkan untuk melakukan uji parsial jika diperoleh keputusan tolak H_0 pada uji simultan, serta menghitung koefisien determinasi (R^2).
9. Menginterpretasikan model yang didapatkan dan membuat kesimpulan.

3. HASIL DAN PEMBAHASAN

Analisis Deskriptif

Deskripsi data dinyatakan dalam statistika deskriptif meliputi rata-rata, standar deviasi, nilai minimum dan nilai maksimum.

Tabel 2 Statistika Deskriptif Data Penelitian

Variabel	Mean	Standar Deviasi	Minimum	Maksimum
y	660,08	570,22	65,00	2708
x_1	6,52	2,62	2,42	12,38
x_2	8,11	4,47	2,12	26,92
x_3	77,69	14,32	47,50	100,00
x_4	15,96	5,62	4,30	26,90

Berdasarkan tabel 2, dapat dilihat bahwa secara rata-rata, terdapat 660,08 kasus tuberculosis di kabupaten/kota di Provinsi Kalimantan Barat, Kalimantan Selatan, dan Kalimantan Timur, di mana jumlah kasus tuberculosis terendah yang ditemukan di kabupaten/kota di ketiga provinsi ini adalah sebesar 65 di Kabupaten Mahakam Ulu. Adapun kasus *tuberculosis* terbanyak ditemukan di Kota Banjarmasin dengan 2708 kasus *tuberculosis*. Selain itu, dapat kita lihat bahwa persentase penduduk miskin secara rata-rata di ketiga daerah ini adalah sebesar 6,52 persen dengan persentase terendah adalah 2,42 persen di Kota Balikpapan dan yang tertinggi adalah 12,38 persen di Kota Melawi. Adapun jika ditinjau dari persentase tenaga kerja kesehatan terlatih, terlihat bahwa Kota Banjarmasin mempunyai persentase tenaga kerja kesehatan terlatih yang tertinggi dengan 26,92% tenaga kerja kesehatan yang terlatih. Adapun yang terendah adalah 2,12 persen di Kabupaten Mahakam Ulu. Selanjutnya, dapat dilihat bahwa persentase keluarga yang memperoleh sanitasi layak tertinggi adalah kabupaten Mahakam Ulu, yang menandakan bahwa kurang dari setengah penduduk di kabupaten ini mendapatkan akses sanitasi yang layak. Adapun yang tertinggi adalah Kota Samarinda dan Kota Balikpapan. Selanjutnya, jika ditinjau dari persentase gizi buruk di masyarakat, terlihat bahwa Kabupaten Tanah Bumbu mempunyai persentase terendah, yaitu sebesar 4,3% yang menandakan gizi masyarakat di kabupaten ini paling baik dibandingkan dengan kabupaten/kota lain di ketiga provinsi ini. Adapun persentase tertinggi terdapat di Kabupaten Kapuas Hulu yang menandakan bahwa di kabupaten ini kualitas gizi yang didapat relatif buruk dikarenakan masih terdapat lebih dari seperempat masyarakatnya memperoleh gizi yang masih relatif buruk.

Model Regresi Nonparametrik Multivariabel dengan Pendekatan *Spline Truncated*

Pemilihan model terbaik dilakukan dengan membandingkan nilai dari GCV yang paling minimum.

Tabel 3 Nilai GCV Minimum Setiap Pemilihan Titik Knot Optimal

Banyak Titik Knot	Nilai GCV Minimum	R ²
Satu Titik Knot	83.740,76	84,84%
Dua Titik Knot	53.981,90	92,41%
Tiga Titik Knot	40.813,24	95,84%
Kombinasi Titik Knot (1,2,2,2)	83.740,76	92,33%

Tabel 3. Menunjukkan bahwa pemodelan yang menghasilkan nilai GCV paling minimum terdapat pada pemodelan regresi nonparametrik *spline truncated* dengan menggunakan tiga titik knot dengan nilai R² sebesar 95,84%. Oleh karena itu, diputuskan bahwa model terbaik

yang akan dipilih adalah model regresi nonparametrik *spline truncated* dengan menggunakan tiga titik knot. Berikut akan ditunjukkan nilai titik knot untuk 3 titik knot pada tiap variabel:

Tabel 4 Nilai GCV untuk 3 Titik Knot

No	x_1	x_2	x_3	x_4	GCV
1	8,92	14,65	81,79	19,06	46.445,34
	11,57	19,74	95,71	25,06	
	11,77	20,14	96,79	25,52	
2	9,13	15,04	82,86	19,52	46.140,15
	9,74	16,22	86,07	20,90	
	9,94	16,61	87,14	21,37	
3	9,33	15,44	83,93	19,98	43.153,43
	9,53	15,83	85,00	20,44	
	9,74	16,22	86,07	20,90	
4	9,33	15,44	83,93	19,98	40.813,24
	9,53	15,83	85,00	20,44	
	10,14	17,00	88,21	21,83	
5	9,33	15,44	83,93	19,98	42.919,37
	9,53	15,83	85,00	20,44	
	12,18	20,92	98,93	26,44	

Diperoleh dari Tabel 4 nilai GCV minimum untuk model regresi nonparametrik *spline truncated* dengan tiga knot adalah sebesar 40.813,24. Nilai ini diperoleh dari tiga titik knot optimal pada setiap variabel. Tiga titik knot pada variabel persentase penduduk miskin (x_1) berada pada titik knot 9,33; 9,53 dan 10,14; variabel prediktor persentase tenaga kesehatan terlatih (x_2) berada pada titik knot 15,44; 15,83 dan 17,00; variabel persentase keluarga dengan akses terhadap sanitasi yang layak (x_3) berada pada titik knot 83,93; 85,00 dan 88,21; serta variabel persentase gizi buruk masyarakat (x_4) berada pada titik knot 19,98; 20,44 dan 21,83. Jika dibentuk kedalam persamaan hasil dari estimasi parameter dengan menggunakan 3 titik knot dapat dilihat pada persamaan (5):

$$\begin{aligned}
 y_i = & -393,02 + 13,07 + 978,56(x_1 - 9,33)_+ - 1988,51(x_1 - 9,53)_+ + 1251,88(x_1 - 10,14)_+ \\
 & + 77,17 + 438,08(x_2 - 15,44)_+ + 135,66(x_2 - 15,83)_+ - 771,61(x_2 - 17,00)_+ - 2,63 + 1256,55(x_3 - 83,93)_+ \\
 & - 1779,70(x_3 - 85,00)_+ + 585,48(x_3 - 88,21)_+ + 37,55 - 2329,82(x_4 - 19,98)_+ + 2941,69(x_4 - 20,44)_+ - 725,67(x_4 - 21,83)_+
 \end{aligned}
 \tag{5}$$

Interpretasi Model

1. Interpretasi model untuk variabel x_1 , apabila variabel x_2 , x_3 dan x_4 dianggap konstan, maka besar pengaruh persentase penduduk miskin terhadap jumlah kasus *tuberculosis* adalah:

$$\hat{y}_i = 13,07x_1 + 978,56(x_1 - 9,33)_+ - 1988,51(x_1 - 9,53)_+ + 1251,88(x_1 - 10,14)_+ \tag{6}$$

$$\hat{y} = \begin{cases} 13,07x_1 & x_1 < 9,33 \\ 991,63x_1 - 9129,98 & 9,33 \leq x_1 < 9,53 \\ -996,88x_1 + 18950,60 & 9,53 \leq x_1 < 10,14 \\ 255,00x_1 + 12694,06 & x_1 \geq 10,14 \end{cases}$$

Model diatas dapat diinterpretasikan jika persentase penduduk miskin berada diatas angka 10,14 persen dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik sampai angka 255 kasus.

2. Interpretasi model untuk variabel x_2 , apabila variabel x_1 , x_3 dan x_4 dianggap konstan, maka besar pengaruh persentase tenaga kesehatan terlatih terhadap jumlah kasus *tuberculosis* adalah

$$\hat{y}_i = 77,17x_2 + 438,08(x_2 - 15,44)_+ + 135,66(x_2 - 15,83)_+ - 771,61(x_2 - 17,00)_+ \quad (7)$$

$$\hat{y} = \begin{cases} 77,17x_2 & x_2 < 15,44 \\ 515,25x_2 - 6763,95 & 15,44 \leq x_2 < 15,83 \\ 650,91x_2 - 2147,50 & 15,83 \leq x_2 < 17,00 \\ -120,70x_2 + 13117,37 & x_2 \geq 17,00 \end{cases}$$

Model diatas dapat diinterpretasikan jika persentase tenaga kesehatan terlatih berada dibawah angka 15,44 persen dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik sampai angka 77 kasus. Hal ini menunjukkan bahwa semakin tinggi persentase tenaga kesehatan terlatih, semakin tinggi pula tingkat skrining *tuberculosis*, dan menyebabkan kasus *tuberculosis* yang diketahui menjadi meningkat.

3. Interpretasi model untuk variabel x_3 , apabila variabel x_1 , x_2 dan x_4 dianggap konstan, maka besar pengaruh persentase keluarga dengan akses terhadap sanitasi yang layak terhadap jumlah kasus *tuberculosis* adalah

$$\hat{y}_i = -2,63x_3 + 1256,55(x_3 - 83,93)_+ - 1779,70(x_3 - 85,00)_+ + 585,48(x_3 - 88,21)_+ \quad (8)$$

$$\hat{y} = \begin{cases} -2,63x_3 & x_3 < 83,93 \\ 1253,92x_3 - 105462,24 & 83,93 \leq x_3 < 85,00 \\ -525,78x_3 + 151274,50 & 85,00 \leq x_3 < 88,21 \\ 59,70x_3 + 51645,19 & x_3 \geq 88,21 \end{cases}$$

Model diatas dapat diinterpretasikan jika persentase keluarga dengan akses terhadap sanitasi yang layak berada diantara angka 83,93% hingga 85,00% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik sampai angka 1253 kasus. Selanjutnya jika keluarga dengan akses terhadap sanitasi yang layak berada diantara 85,00% smpa 88,21% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung turun sampai angka 525 kasus. Sedangkan jika persentase keluarga dengan akses terhadap sanitasi yang layak berada diatas 88,21% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik hingga mencapai angka 59 kasus.

4. Interpretasi model untuk variabel x_4 , apabila variabel x_1 , x_2 dan x_3 dianggap konstan, maka besar pengaruh persentase gizi buruk masyarakat terhadap jumlah kasus *tuberculosis* adalah

$$\hat{y}_i = 37,55x_4 - 2329,82(x_4 - 19,98)_+ + 2941,69(x_4 - 20,44)_+ - 725,67(x_4 - 21,83)_+ \quad (9)$$

$$\hat{y} = \begin{cases} 37,55x_4 & x_4 < 19,98 \\ -2291,45x_4 + 46533,42 & 19,98 \leq x_4 < 20,44 \\ 650,24x_4 - 60128,14 & 20,44 \leq x_4 < 21,83 \\ -75,43x_4 + 1584137,61 & x_4 \geq 21,83 \end{cases}$$

Model diatas dapat diinterpretasikan jika persentase gizi buruk masyarakat berada dibawah angka 19,98% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik sampai angka 375 kasus. Kemudian apabila persentase gizi buruk masyarakat diantara 19,98% hingga 20,44% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung turun sampai angka 2291 kasus. Sedangkan jika persentase gizi buruk masyarakat berada diantara angka 20,44% sampai 21,83% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung naik hingga mencapai angka 650 kasus. Selanjutnya jika persentase gizi buruk masyarakat berada diatas angka 21,83% dan mengalami 1 persen kenaikan maka jumlah kasus *tuberculosis* cenderung turun hingga mencapai angka 75 kasus.

4. SIMPULAN DAN SARAN

Simpulan

Berdasarkan pengujian signifikansi parameter diperoleh bahwa semua variabel berpengaruh secara signifikan terhadap jumlah kasus *tuberculosis* karena sudah mempunyai nilai p-value yang kurang dari taraf uji yang ditetapkan. Keempat variabel tersebut diantaranya adalah persentase penduduk miskin, persentase tenaga kesehatan terlatih, persentase keluarga memperoleh sanitasi dan persentase gizi buruk masyarakat. Model tersebut memiliki nilai *R-Squared* sebesar 95,84 persen, yang berarti keempat variabel ini sudah cukup menjelaskan 95,84 persen variasi jumlah kasus *tuberculosis* di kabupaten/kota di ketiga provinsi ini. Adapun 4,16 persen variasi lainnya dijelaskan oleh variabel lain yang tidak berada di dalam model regresi.

Saran

Pada penelitian selanjutnya yaitu dapat menggunakan metode regresi nonparametrik *spline truncated* dengan pemilihan metode pemilihan titik knot optimal menggunakan metode CV (*Cross Validation*).

DAFTAR PUSTAKA

- Eubank, R.L. (1988). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker Inc.
- Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing 2nd*. New York: Marcel Dekker Inc.
- Kementerian Kesehatan Republik Indonesia. (2015). *Survei Prevelensi Tuberkolusis 2013-2014*. Jakarta: Kementerian Kesehatan RI.
- Putra, I.M.B., Srinadi, I.G.A.M., Sumarjaya, I.W. (2015). "Pemodelan Regresi *Spline* (Studi Kasus: Herpindo Jaya Cabang Ngaliyan)". *Jurnal Matematika*. 04(03),110-114.
- Wahba, G. (1990). *Spline Models for Observation Data*. *Society for Industrial and Applied Mathematics*. Philadelphia: Pennsylvania.