

PEMODELAN ANGKA HARAPAN HIDUP SUMATERA UTARA MENGGUNAKAN REGRESI SEMIPARAMETRIK *SPLINE*

Melati Puspita Sari Lubis¹, Rina Widyasari², Rima Aprilia³

Universitas Islam Negeri Sumatera Utara ^{1,2,3}

melatipuspita11@gmail.com¹, rina_widyasari@uinsu.ac.id², rima_apriliah@uinsu.ac.id³

Corresponding Author: Melati Puspita Sari Lubis email: melatipuspita11@gmail.com

Abstrak. Angka harapan hidup merupakan dugaan hidup rata-rata seseorang atau dengan kata lain umur yang mungkin didapat seseorang yang lahir pada waktu tertentu. Menurut data Badan Pusat Statistik, Sumatera Utara berada pada posisi terendah di pulau Sumatera dengan rata-rata angka harapan hidup dengan angka 67,22 pada laki-laki dan 71,08 pada perempuan. Angka ini juga masih dibawah rata-rata angka harapan hidup nasional. Hal ini menunjukkan program-program kesehatan di Sumatera Utara perlu menjadi prioritas dalam pembangunan manusia di Sumatera Utara. Metode yang digunakan untuk memodelkan angka harapan hidup Sumatera Utara yaitu Regresi Semiparametrik Spline *Truncated*. Regresi semi parametrik merupakan perpaduan antara regresi parametrik dan regresi nonparametrik dengan sebagian bentuk pola data yang diketahui dan ada sebagian pola data yang tidak diketahui. Adapun tujuan yang didapatkan untuk mengetahui model terbaik dari hasil pola data hubungan variabel respon dan variabel prediktor. Hasil penelitian ini diperoleh nilai MSE sebesar 7,3893 dan GCV sebesar 1,3169. Dengan persamaan model sebagai berikut:

$$\hat{y} = -49,0340 + 0,2626x_1 - 1,4231x_2 - 1,2746(x_2 - 183,54)_+ + 3,9979x_3 - 41,7015(x_3 - 11,39)_+ + 1,1202x_4 - 1,2015(x_4 - 88,95)_+$$

Dengan nilai koefisien determinasi sebesar 39,47% yang dapat diartikan kemampuan variabel prediktor dalam menjelaskan variabel responnya sebesar 39,47% dan sisanya dipengaruhi oleh faktor-faktor lain.

Kata Kunci: Angka Harapan Hidup, Regresi Semiparametrik, Spline *Truncated*.

Abstract. Life expectancy is an estimate of a person's average life or in other words the age that someone born at a certain time might get. According to the Central Bureau of Statistics, North Sumatra is at a low on the island of Sumatra with an average life expectancy of 67,22 in men and 71,08 in women. This figure is also still below the national average life expectancy. This shows that health programs in North Sumatra need to be a priority in human development in North Sumatra. The method used to model North Sumatera's life expectancy is Spline *Truncated* Semiparametric Spline *Truncated* Regression. Semiparametric regression is a combination of parametric regression and nonparametric regression with some forms of known data patterns and some unknown data patterns. The purpose obtained is to find out the best model of the results of the data patterns of response variable relationships and predictor variables. The results of this study obtained the MSE value of 7,3893 and GCV of 1,3169. With the model equation as follows:

$$\hat{y} = -49,0340 + 0,2626x_1 - 1,4231x_2 - 1,2746(x_2 - 183,54)_+ + 3,9979x_3 - 41,7015(x_3 - 11,39)_+ + 1,1202x_4 - 1,2015(x_4 - 88,95)_+$$

With a determination coefficient value of 39,47% which can be interpreted as the ability of the predictor variable in explaining its response variables by 39,47% and the rest is influenced by other factors.

Keywords: Life Expectancy, Semiparametric Regression, *Truncated Spline*.

A. Pendahuluan

Hidup sehat merupakan salah satu indikator yang diperlukan dalam Indeks Pembangunan Manusia (IPM). IPM bermakna pengukuran terhadap harapan hidup sehat, pendidikan dan standar pendapatan hidup manusia. *United Nation Development Programme* (UNDP) mempunyai tiga dimensi dasar yang menjadi indikator dalam indeks pembangunan manusia dan telah disetujui dunia yaitu dimensi berumur panjang dan hidup sehat, memperoleh pendidikan dan memiliki akses terhadap sumber-sumber kebutuhan agar hidup secara layak (BPS



Kabupaten Simalungun, 2019). Indikator pembangunan manusia pada dimensi berumur panjang dan hidup sehat adalah angka harapan hidup.

Angka harapan hidup pada suatu umur tertentu dimaknai dengan rata-rata tahun hidup yang akan dilalui oleh seseorang yang telah berjaya mencapai umur tersebut dalam situasi yang berlangsung di lingkungan masyarakatnya. Angka harapan hidup disuatu wilayah atau negara bisa mengalami naik turun disetiap tahunnya. Bahkan angka harapan hidup bisa berbeda-beda di setiap daerah disatu Negara. Berdasarkan data Badan Pusat Statistik (BPS) pada tahun 2020, pada pulau Sumatera dengan sepuluh provinsi Riau menduduki posisi teratas pada angka 69,75 untuk laki-laki dan 73,55 untuk perempuan. Sumatera Utara berada pada posisi terendah di pulau Sumatera dengan rata-rata angka harapan hidup dengan angka 67,22 pada laki-laki dan 71,08 pada perempuan. Angka ini juga masih dibawah rata-rata angka harapan hidup nasional yaitu 69.59 tahun pada laki-laki dan 73.46 tahun pada perempuan. Hal ini menunjukkan program-program kesehatan di Sumatera Utara perlu menjadi prioritas dalam pembangunan manusia di Sumatera Utara.

Metode yang dipakai adalah regresi semiparametrik. Regresi semiparametrik adalah perpaduan dari regresi parametrik dan regresi nonparametrik (Adawiyah, 2018). Regresi parametrik cenderung memiliki bentuk pola hubungan mendekati garis linier dan membentuk garis lurus sedangkan regresi nonparametrik kecenderungan tidak memiliki hubungan pola tertentu. Regresi semiparametrik menghasilkan pola hubungan antara variabel respon dengan variabel prediktor yang mencakup regresi parametrik ($x_i\beta$) dan regresi nonparametrik $f(x_i)$ yang kemudian menghasilkan model terbaik dari pola hubungan antara variabel respon dengan variabel prediktor tersebut (Side, 2020).

Salah satu model regresi semiparametrik yang sering dipakai yaitu *spline*. *Spline* merupakan potongan polynomial tersegmen yang mempunyai sifat fleksibilitas. *Spline* juga diinterpretasikan dengan model optimal terputus tetapi masih bersifat berkesinambungan pada titik knotnya. Titik perpaduan dari potongan-potongan tersebut atau titik yang menyatakan terbentuknya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut knot (Adawiyah, 2018). Salah satu basis fungsi yang sering dipakai terhadap pendekatan *spline*, yaitu *spline truncated*. *Spline truncated* dapat menangani data yang mempunyai perilaku yang berubah-ubah pada sub-sub interval tertentu dan cenderung memilih sendiri estimasi data kemanapun pola data bergerak. Kelebihan dari *spline truncated* ini berlaku jika dalam fungsi terdapat adanya titik-titik knot (Dani, 2021).

Adapun variabel yang digunakan yaitu BPJS kesehatan, penduduk miskin, rata-rata lama sekolah dan TPAK. Dengan begitu, akan dilakukan pemodelan pada angka harapan hidup Sumatera Utara dengan regresi semi parametrik *spline*.

B. Metodologi Penelitian

Penelitian ini menggunakan penelitian kuantitatif dengan menggunakan metode regresi semiparametrik *spline truncated* dimana metode ini merupakan gabungan dari regresi parametrik dan regresi nonparametrik. Jenis data yang digunakan pada penelitian ini yaitu data sekunder yang diperoleh dari Badan Pusat Statistik Sumatera Utara. Data yang digunakan dalam penelitian ini adalah data angka harapan hidup, BPJS Kesehatan, Penduduk Miskin, Rata-Rata Lama Sekolah, Tingkat Partisipasi Angkatan Kerja tahun 2020.

Prosedur penelitian yang dilaksanakan guna mencapai tujuan penelitian yaitu:

1. Mengumpulkan data dari Badan Pusat Statistik Provinsi Sumatera Utara.
2. Menentukan variabel respon dan variabel prediktor yang akan digunakan.
3. Membuat *scatter plot* variabel respon dengan masing-masing variabel prediktor.
4. Memodelkan data angka harapan hidup dengan variabel prediktor yang telah ditetapkan dengan regresi semi parametrik *spline*.



5. Memilih titik knot optimal dengan menggunakan *Generalized Cross Validation* (GCV) yang paling kecil.
6. Menghitung nilai koefisien determinasi (R^2).
7. Melakukan interpretasi model dan menarik kesimpulan.

C. Hasil Penelitian dan Pembahasan

1. Karakteristik Angka Harapan Hidup Serta Faktor Yang Mempengaruhi

Data yang digunakan dalam penelitian ini untuk membentuk model yaitu dengan data yang diambil dari Bdn Pust Statistik tahun 2020. Variabel-variabel yang digunakan dalam penelitian ini adalah Angka harapan hidup (Y) dengan BPJS kesehatan (X_1), Penduduk miskin (X_2), Rata-rata lama sekolah (X_3), Tingkat Partisipasi Angkatan Kerja (X_4). Karakteristik angka harapan hidup serta faktor yang mempengaruhi yaitu nilai rata-rata, varians, standar deviasi, nilai minimum dan nilai maksimum, dapat disajikan dalam table 1.

Tabel 1 Karakteristik angka harpan hidup serta faktor yang mempengaruhi

Variabel	Rata-rata	Varians	S. Deviasi	Min	Max
Y	69,023333	6,272247	2,504445	62,6	73,55
x_1	20,26	76,66521875	8,640903	4,34	36,66
x_2	38,887	1157,875461	34,027569	4,59	183,54
x_3	9	2,043499	1,429510	5,36	11,39
x_4	71	77,357304	8,795300	51,83	88,95

Kemudian dalam melakukan penelitian untuk memodelkan angka harapan hidup menggunakan pendekatan regresi semiparametrik spline, dapat dilakukan dengan cara mencari nilai koefisien korelasinya agar mengetahui apakah hubungan linier nya lemah atau tidak dengan menggunakan rumus.

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (1)$$

$$r_{x_1y} = \frac{1536282,442 - 1522871,467}{\sqrt{[1536282,442 - 446999,2164][6623,4938]}} = \frac{13410,9756}{84940,33607} = 0,158$$

$$r_{x_2y} = \frac{2945733,295 - 2923062,241}{\sqrt{[2945733,295 - 1646858,89][6623,4938]}} = \frac{22671,0539}{92752,82512} = 0,244$$

$$r_{x_3y} = \frac{685136,4927 - 683968,7756}{\sqrt{[92326,014 - 90168,0784][6623,4938]}} = \frac{1167,7171}{3780,61808} = 0,308$$

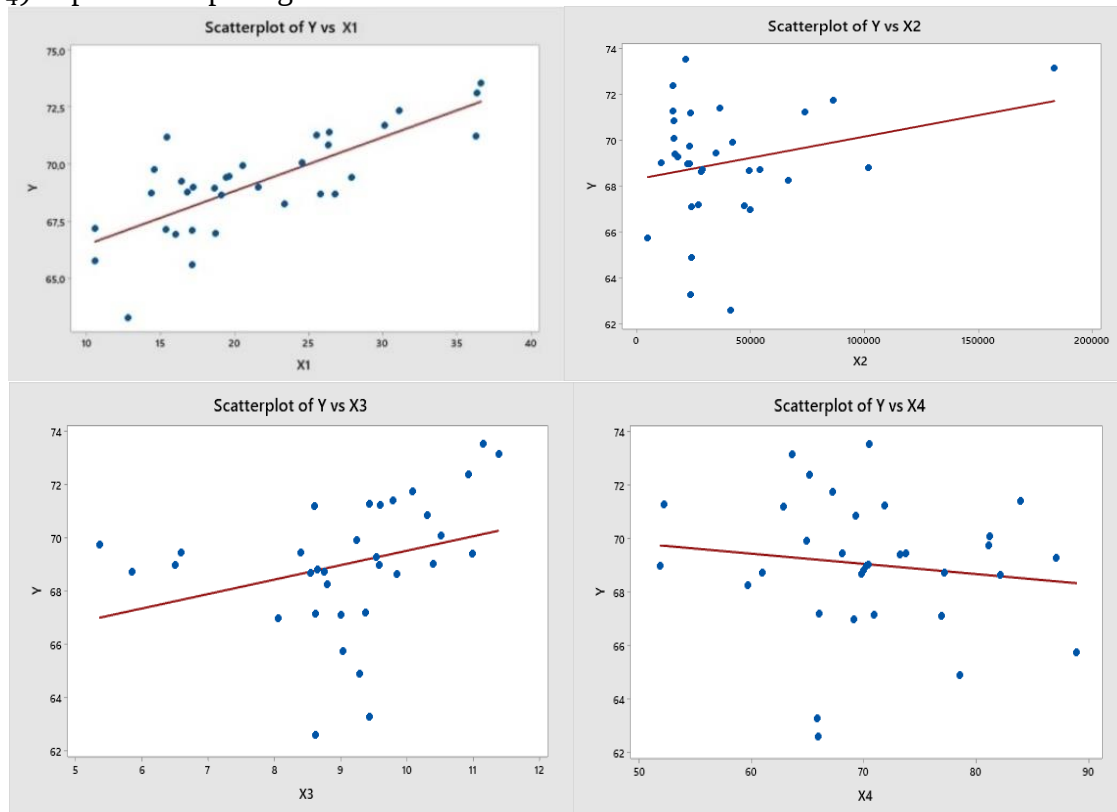
$$r_{x_4y} = \frac{5304548,792 - 5307659,654}{\sqrt{[5511521,354 - 5429832,04][6623,4938]}} = \frac{-3110,8616}{23260,88269} = -0,133$$

Tabel 2 Nilai Korelasi Antara Variabel

Korelasi	Koefisien Korelasi
y terhadap X_1	0,158
y terhadap X_2	0,244
y terhadap X_3	0,308
y terhadap X_4	-0,133

Dapat dilihat pada nilai koefisien y terhadap X_3 lebih besar diantara keempat variabel yang lainnya, hal tersebut mengartikan jika hubungan variabel tersebut lebih kuat dibandingkan dengan variabel lainnya

Langkah berikutnya untuk melihat pola hubungan antara Angka harapan hidup (Y) dengan BPJS kesehatan (X_1), Penduduk miskin (X_2), Rata-rata lama sekolah (X_3), TPAK (X_4) dapat dilihat pada gambar 1 berikut.



Gambar 1. Scatterplot angka harapan hidup terhadap masing-masing variable predictor

Pada gambar 1. dapat dilihat bahwa X_1 terlihat ada suatu pola hubungan linear garis lurus mengarah ke kanan atas yang membentuk pola data yang sama dan mendekati garis linear, maka variabel tersebut merupakan variabel yang bersifat parametrik, sedangkan, X_2 , X_3 , dan X_4 pola hubungannya tidak terdapat kecenderungan hubungan pola tertentu yang diartikan hubungannya bersifat non parametrik.

2. Pemilihan Titik Knot

Pemilihan titik knot optimal dihasilkan dengan menggunakan bantuan software Matlab R2017b yang telah disesuaikan dengan metode regresi spline *Truncated* yang digunakan dalam penelitian kali ini. Selanjutnya dipilih nilai GCV terendah yang didapatkan pada hasil perhitungannya. GCV yang dihasilkan dengan memakai satu titik knot disajikan di table 3 Titik knot yang dihasilkan merupakan titik yang diketahui menjadi letak perubahan pada data penelitian ini.

Tabel 3. Nilai GCV Dan MSE Dengan 1 Titik Knot

Titik Knot			MSE	GCV
k_1	k_2	k_3		
183.54	11.39	88.95	7.3893	1.3169
16.32	10.31	69.27	1.8217	2.4218
49.18	8.54	69.82	1.4806	2.2895
34.75	6.58	73.71	2.1685	2.0937

Berdasarkan Tabel 3. dapat dilihat bahwa model spline *Truncated* terbaik dengan satu titik knot adalah model dengan nilai GCV terkecil yaitu 1,3169 dan nilai MSE sebesar 7,3893. Model ini memiliki titik knot pada $k_{21}=183.54$, $k_{31}=11.39$, dan $k_{41}=88.95$.

3. Pemodelan Regresi Semiparametrik Spline Truncated

Pemodelan dengan y adalah variabel respon dan x adalah variable, bentuk umum fungsi spline *Truncated* dengan K titik knot, $k_1, k_2, \dots, k_k$ dan satu prediktor dapat disajikan sebagai berikut.

$$y_i = \sum_{j=1}^q \beta_{ji} x_i + \sum_{j=1}^k \alpha_{ji} (x_i - k_j)_+ \quad (2)$$

Dengan fungsi *Truncated*

$$(x_i - k_j)_+ = \begin{cases} (x_i - k_j) & ; x_i \geq k_j \\ 0 & ; x_i < k_j \end{cases} \quad (3)$$

Tabel 4 Estimasi Model Regresi Spline dengan 1 Titik Knot

Parameter	Estimasi Parameter
$\hat{\beta}_0$	-49,0340
$\hat{\beta}_1$	0,2626
$\hat{\beta}_2$	-1,4231
$\hat{\beta}_3$	3,9979
$\hat{\beta}_4$	1,1202
γ_1	-1,2746
γ_2	-41,7015
γ_3	-1,2015

Berdasarkan Tabel 4. estimasi parameter model terbaik pada titik knot $k_{21} = 183,45$, $k_{31} = 11,39$, dan $k_{41} = 88,95$ menghasilkan model regresi spline *Truncated* untuk variabel y dapat disajikan sebagai berikut:

$$\hat{y} = -49,0340 + 0,2626x_1 - 1,4231x_2 - 1,2746(x_2 - 183,54)_+ + 3,9979x_3 - 41,7015(x_3 - 11,39)_+ + 1,1202x_4 - 1,2015(x_4 - 88,95)_+$$

$$\hat{y} = -49,0340 + f_1(x_1) + f_2(x_2 - k) + f_3(x_3 - k) + f_4(x_4 - k)$$

Dengan

$$\hat{f}(x_1) = 0,2626x_{1+}$$

$$\hat{f}(x_2 - k) = -1,4231x_2 - 1,2746(x_2 - 183,54)_+$$

$$\hat{f}(x_3 - k) = 3,9979x_3 - 41,7015(x_3 - 11,39)_+$$

$$\hat{f}(x_4 - k) = 1,1202x_4 - 1,2015(x_4 - 88,95)_+$$

Adapun persamaan model spline *Truncated* terbaik untuk y adalah:

$$\hat{y}(x_2) = \begin{cases} -1,4231x_2, & x_2 < 183,54 \\ 233,9400 - 2,6977x_2, & x_2 \geq 183,54 \end{cases}$$

$$\hat{y}(x_3) = \begin{cases} 3,9979x_3, & x_3 < 11,39 \\ 474,9800 - 37,7036x_3, & x_3 \geq 11,39 \end{cases}$$

$$\hat{y}(x_4) = \begin{cases} 1,1202x_4, & x_4 < 88,95 \\ 106,8734 - 0,0813x_4, & x_4 \geq 88,95 \end{cases}$$

4. Uji Kelayakan Model

Setelah melaksanakan pemodelan regresi spline *Truncated* birespon dengan 1 titik knot maka selanjutnya dilakukan uji kelayakan model memakai (R^2) yang bertujuan untuk mengetahui besarnya pengaruh variabel respon terhadap variabel predictor.

Tabel 5 Nilai Koefisien Determinasi

S	R-sq	R-sq(adj)	R-sq(pred)
2,08307	39,47%	30,82%	11,71%

Dengan nilai koefisien determinasi sebesar 39,47% yang dapat diartikan kemampuan variabel prediktor dalam menjelaskan variabel responnya sebesar 39,47% dan sisanya dipengaruhi oleh faktor-faktor lain.

D. Kesimpulan

Angka harapan hidup digunakan sebagai indikator untuk mengukur kesehatan individu di suatu daerah. Menurut data Badan Pusat Statistik Sumatera Utara berada pada posisi terendah di pulau Sumatera dengan rata-rata angka harapan hidup dengan angka 67,22 pada laki-laki dan 71,08 pada perempuan. Angka ini juga masih dibawah rata-rata angka harapan hidup nasional. Hal ini menunjukkan program-program kesehatan di Sumatera Utara perlu menjadi prioritas dalam pembangunan manusia di Sumatera Utara.

Metode semiparametrik *spline* ini dipakai pada angka harapan hidup dengan pola data yang bersifat parametrik terdapat pada variabel BPJS kesehatan sementara data yang bersifat nonparametrik terdapat pada variabel penduduk miskin, rata-rata lama sekolah dan TPAK. Diperoleh model regresi semiparametrik spline *truncated* terbaik pada titik knot optimal $k_{21} = 183,54$, $k_{31} = 11,39$, dan $k_{41} = 88,95$ dengan nilai MSE sebesar 7,3893 dan GCV sebesar 1,3169. Persamaan model spline *Truncated* sebagai berikut:

$$\hat{y} = -49,0340 + 0,2626x_1 - 1,4231x_2 - 1,2746(x_2 - 183,54)_+ + 3,9979x_3 - 41,7015(x_3 - 11,39)_+ + 1,1202x_4 - 1,2015(x_4 - 88,95)_+$$

Dengan nilai koefisien determinasi sebesar 39,47% yang dapat diartikan kemampuan variabel prediktor dalam menjelaskan variabel responnya sebesar 39,47% dengan 60,53% sisanya dijelaskan oleh faktor lain yang tidak diketahui diluar model.

DAFTAR PUSTAKA

- Adawiyah, R.dkk. (2018). Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar). *Journal Of Mathematics, Computations and Statistics*. 2 (1). 70-81.
- Badan Pusat Statistik Sumatera Utara. (2020). *Indeks Pembangunan Manusia Provinsi Sumatera Utara Tahun 2020*. CV. Rabbani. Medan.
- Dani, A. T. R.dkk. (2021). Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal (Studi Kasus: Persentase Penduduk Miskin Provinsi Papua Tahun 2016-2019). *Inferensi*. 4 (1), 47-55.



- Gabrela, P.P.dkk. (2019). Pemodelan Angka Harapan Hidup di Provinsi Papua Menggunakan Pendekatan Regresi Nonparametrik Spline Truncated. *Jurnal Sains Dan Seni ITS*. 8 (2), 341-348.
- Laome, L. (2015). Model Regresi Semiparametrik Spline Untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV. *Jurnal Matematika Murni dan Terapan*, 13 (2), 189-194.
- Muda, R. dkk. (2019). Pengaruh Angka Harapan Hidup, Tingkat Pendidikan dan Pengeluaran Perkapita Terhadap Pertumbuhan Ekonomi di Sulawesi Utara. *Jurnal Berkala Ilmiah Efisiensi*. 19 (1), 44-55.
- Nasution, Y., dan Hasibuan, L. S. (2018). Analisis Pengaruh Belanja Sektor Kesehatan Terhadap Angka Harapan Hidup Di Sumatera Utara. *Ekonomikawan : Jurnal Ilmu Ekonomi Dan Studi Pembangunan*. 18 (1), 70-92.
- Nisa Khaerun. (2020). Modeling East Java Indonesia Life Expectancy Using Semiparametric Regression Mixed Spline Truncated And Fourier Series. *Media Statistika*. 13 (2), 149-160.
- Pratiwi, D. A. dkk. (2017). Pendekatan Regresi Semiparametrik Spline untuk Memodelkan Rata-Rata Umur Kawin Pertama (UKP) di Provinsi Jawa Timur. *Jurnal Sains Dan Seni ITS*. 6 (1), 129-136.
- Side, S. dkk. (2020). *Model Regresi Semiparametrik Spline Untuk Data Longitudinal Pada Kasus Penderita Demam Berdarah Dengue Di Kota Makassar*. *Journal of Mathematics, Computation and Statistics*. 3 (1), 20-31.
- Utami, Tiani Wahyu. (2018). Pendekatan Regresi Semi Parametrik Spline Truncated Untuk Pemodelan Tingkat Pengangguran Terbuka Di Jawa Tengah. *Statistika*. 6 (2), 172-176.
- Yani, N. W. M. N. dkk. (2017). Aplikasi Model Regresi Semiparametrik Spline Truncated (Studi Kasus: Pasien Demam Berdarah Dengue (DBD) Di Rumah Sakit Puri Raharja). *E-Jurnal Matematika*. 6 (1), 65-73.

