

# Pemodelan Kemiskinan di Sumatera Utara Menggunakan Regresi Nonparametrik Kernel dan Splines

(*Kernel and Splines Nonparametric Regression for Poverty Model in North Sumatera*)

Hasrat Ifolala Zebua

Program Studi Magister Statistika Terapan, Fakultas MIPA, Univeristas Padjadjaran, Indonesia  
Gedung 2 Lt.5 Kampus UNPAD Jl. Dipati Ukur No. 35 Bandung  
E-mail: hasrat20001@mail.unpad.ac.id

## ABSTRAK

Penanggulangan kemiskinan merupakan tujuan utama dari *Sustainable Development Goals* (SDGs). Masyarakat miskin disebabkan oleh rendahnya modal manusia. Salah satu indikator yang dapat mengukur modal manusia adalah Indeks Pembangunan Manusia (IPM). Sumatera Utara masih menempati posisi ke-lima dengan tingkat kemiskinan tertinggi di Pulau Sumatera, padahal memiliki jumlah penduduk terbanyak di Pulau Sumatera. Pola hubungan antara kemiskinan dan IPM berdasarkan penelitian sebelumnya masih belum dapat ditentukan secara jelas karena memiliki hasil yang berbeda-beda sehingga pada penelitian ini digunakan pemodelan regresi nonparametrik karena sifatnya yang fleksibel dalam mengikuti pola hubungan data serta dapat menghindari kesalahan prespesifikasi model. Regresi nonparametrik yang digunakan pada penelitian ini adalah regresi kernel, *smoothing splines* dan regresi kuantil *smoothing splines*. Regresi kuantil digunakan untuk mengakomodir perbedaan model pada beberapa kuantil data. Hasil perbandingan model diperoleh bahwa metode regresi kuantil *smoothing splines* merupakan metode terbaik karena memiliki kurva regresi yang lebih mengikuti sebaran hubungan data dan nilai GCV dan RMSE yang lebih rendah.

**Kata kunci:** kemiskinan, nonparametrik, kernel, *smoothing splines*, kuantil

## ABSTRACT

*The Sustainable Development Goals' main goal is to reduce poverty (SDGs). Low human capital is the cause of poverty. The Human Development Index is one indicator that can be used to assess human capital (HDI). Despite having the largest population on the island of Sumatra, North Sumatra continues to have the fifth highest poverty rate. Because the pattern of the relationship between poverty and HDI based on previous research is still unclear because the results are inconsistent, nonparametric regression modeling was used in this study because it is flexible in following the pattern of data relationships and can avoid model prespecific errors. The nonparametric regressions used in this study are kernel regression, smoothing splines and quantile regression smoothing splines. Quantile regression is used to accommodate model differences in several data quantiles. The smoothing splines quantile regression method is the best method, according to the results of the model comparison, because it has a regression curve that follows the distribution of data relationships and lower GCV and RMSE values.*

**Keywords:** *poverty, nonparametric, kernel, smoothing splines, quantile*

## PENDAHULUAN

Penanggulangan kemiskinan merupakan tujuan utama dari *Sustainable Development Goals* (SDGs). Kemiskinan merupakan masalah utama kependudukan yang banyak dihadapi negara-negara di dunia tidak terkecuali di Indonesia. Hal ini sesuai dengan agenda prioritas Presiden Joko Widodo (Nawacita) yang tercantum dalam agenda pembangunan manusia yaitu masalah kemiskinan (Hoelman & dkk, 2015). Berdasarkan Berita Resmi Statistik (BPS, Berita Resmi Statistik Profil Kemiskinan di Indonesia Maret 2021, 2021a) persentase penduduk miskin di Indonesia dari tahun ke tahun terus mengalami penurunan, kecuali tahun 2020 kembali meningkat dan kembali turun pada tahun 2021 yang diakibatkan adanya pandemi covid-19. Pengentasan kemiskinan terus dilakukan oleh pemerintah dengan harapan taraf hidup masyarakat dapat meningkat sehingga angka kemiskinan terus menurun.

Todaro (2006) menyatakan bahwa kemiskinan bukanlah kondisi dimana masyarakat kekurangan suatu komoditi namun lebih cenderung merupakan kondisi dimana masyarakat kurang dapat memaksimalkan fungsi dan mengambil manfaat dari komoditi tersebut. Hal ini tentu saja berkaitan dengan modal manusia (*human capital*). Ini diperkuat oleh Dowling dan Valenzuela (2010) yang menyatakan bahwa masyarakat miskin disebabkan oleh rendahnya modal manusia. Salah satu indikator yang dapat mengukur modal manusia adalah Indeks Pembangunan Manusia (IPM). Konsep pembangunan manusia pertama kali diperkenalkan oleh *United*

*Nations Development Programme* (UNDP). IPM dibentuk berdasarkan 3 dimensi yaitu dimensi umur panjang dan hidup sehat, dimensi pengetahuan, dan dimensi standar hidup layak. Ketiga dimensi ini merupakan pendekatan yang dipilih dalam penggambaran kualitas hidup manusia (BPS, 2021).

Hubungan antara kemiskinan dan pembangunan manusia telah banyak diteliti. Prasetyoningrum dan Sukmawati (2018) dalam hasil penelitiannya menggunakan analisis jalur menyebutkan bahwa IPM berpengaruh signifikan negatif terhadap kemiskinan di Indonesia. Sihite (2020) dengan menggunakan regresi data panel juga menyebutkan bahwa IPM memiliki hubungan negatif dan signifikan terhadap kemiskinan di Provinsi Sumatera Utara. Safuridar dan Putri (2019) mengemukakan hal yang berbeda yaitu dengan regresi data panel IPM berpengaruh positif dan tidak signifikan terhadap kemiskinan di Aceh bagian timur.

Angka kemiskinan di Provinsi Sumatera Utara Maret 2021 sebesar 9,01 persen yang masih lebih kecil dari angka nasional 10,14 persen. Sekalipun demikian, jika dibandingkan dengan provinsi-provinsi di pulau Sumatera, Sumatera Utara masih menempati posisi ke-lima dengan tingkat kemiskinan tertinggi. Hal ini tentu saja masih perlu menjadi perhatian, mengingat Sumatera Utara memiliki jumlah penduduk terbanyak di Pulau Sumatera dan merupakan provinsi ke-empat dengan jumlah penduduk terbanyak se-Indonesia.

Berdasarkan keterangan diatas, maka perlu untuk dilakukan suatu pemodelan kemiskinan dengan faktor yang mempengaruhinya yaitu IPM dengan menggunakan analisis regresi. Pola hubungan antara kemiskinan dan IPM berdasarkan penelitian sebelumnya masih belum dapat ditentukan secara jelas karena memiliki hasil yang berbeda-beda sehingga pada penelitian ini digunakan pemodelan regresi nonparametrik karena sifatnya yang fleksibel dalam mengikuti pola hubungan data. Regresi nonparametrik digunakan untuk menghindari kesalahan prespesifikasi model yang mengakibatkan kurangnya akurasi dari estimator. Pendekatan nonparametrik yang digunakan adalah pendekatan regresi *kernel* dan *smoothing splines*. IPM yang diduga mempengaruhi kemiskinan juga berkemungkinan memiliki pengaruh yang berbeda pada kelompok data tertentu. Perbedaan ini perlu diakomodasi sehingga perlu dibentuk model yang berbeda untuk beberapa kelompok data. Analisis yang dapat digunakan untuk memodelkan data dengan level yang berbeda adalah *Quantile Regression* (QR). Oleh karena itu, dalam penelitian ini perlu dilakukan kombinasi antara regresi nonparametrik dan QR untuk melakukan pemodelan kemiskinan di Provinsi Sumatera Utara.

## METODE

### Kajian Teori

#### *Regresi Nonparametrik*

Regresi nonparametrik merupakan salah satu tata cara statistik yang digunakan untuk mengenali ikatan antara 2 variabel acak ataupun lebih. Regresi nonparametrik dapat disebut juga dengan statistik sebaran bebas. Maksudnya regresi nonparametrik tidak mensyaratkan fungsi serta sebaran dari populasinya (Hardle, 1990) Model umum regresi nonparametrik adalah sebagai berikut:

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

dimana:

$y_i$  = variable respon pengamatan ke-i

$x_i$  = variabel prediktor ke-i

$f(x)$  = fungsi regresi yang tidak diketahui

$e_i$  = residual ke-i, dimana diasumsikan identik, independen, dan berdistribusi normal.

Pada regresi nonparametrik, kurva regresi diasumsikan *smooth* yang berarti berada pada suatu ruang fungsi tertentu sehingga memiliki fleksibilitas (Eubank, 1999). Beberapa model regresi nonparametrik yang sering digunakan yaitu *kernel* dan *spline*.

#### *Regresi Kernel*

Salah satu teknik regresi nonparametrik adalah regresi kernel. Regresi kernel memperkirakan ekspektasi bersyarat dari variabel acak untuk mendapatkan hubungan non-linier antara variabel acak X dan Y, serta untuk mendapatkan estimator dengan bobot yang cocok. Regresi kernel menggunakan penaksir densitas kernel K dengan *bandwidth*  $h$  yang secara umum dapat dituliskan sebagai berikut:

$$K_h = \frac{1}{h} K\left(\frac{x}{h}\right) \quad ; \text{ untuk } -\infty < x < \infty \text{ dan } h > 0 \dots\dots\dots (2)$$

Sehingga penaksir densitas kernel pada  $f(x)$  dapat dituliskan:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \dots\dots\dots (3)$$

Pada estimator kernel, tingkat kemulusan  $\hat{f}_h$  bergantung oleh fungsi kernel K dan *bandwidth* h, namun pengaruh *bandwidth* h lebih dominan dibandingkan fungsi kernel. Nilai h yang kecil membuat kurva kurang mulus, sebaliknya nilai h yang besar akan menghasilkan kurva yang sangat mulus. Untuk itu, perlu dilakukan pemilihan h yang optimal agar kurva yang dihasilkan sesuai. Salah satu pemilihan h yang optimal dengan memilih nilai *Generalized Cross Validation* (GCV) terkecil. Menurut (Hardle, 1990) terdapat beberapa jenis densitas kernel antara lain uniform, triangle, normal (gaussian), epanechnikov, rectangular, dan lain sebagainya. Selain densitas kernel, terdapat juga beberapa estimator untuk menaksir  $f(x)$  antara lain *Nadaraya-Watson Estimator* (NWE), *Gasser-Müller Estimator* (GME), dan *Local Polynomial Estimator* (LPE), namun dalam penelitian ini hanya digunakan NWE.

NWE yang diperkenalkan oleh Nadaraya dan Watson pada tahun 1964 dapat didefinisikan sebagai berikut:

$$\hat{f}(x) = \frac{\sum_{i=1}^n K_h(X_i-x)Y_i}{\sum_{i=1}^n K_h(X_i-x)} \dots\dots\dots(4)$$

Untuk X fixed, penaksir  $\hat{\theta}$  yang meminimumkan

$$\sum_{i=1}^n (Y_i - x)^2 K_h(X_i - x) \dots\dots\dots(5)$$

Memiliki bentuk  $\sum_{i=1}^n a_i Y_i$ . NWE merupakan *minimizer* dari persamaan (5) dimana

$$a_i = \frac{K_h(X_i-x)}{\sum_{i=1}^n K_h(X_i-x)} \dots\dots\dots(6)$$

*Regresi Smoothing Splines*

Teknik regresi nonparametrik yang lain adalah regresi *spline*. Spline punya kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu dan sangat baik dalam menggeneralisasikan pemodelan statistika yang rumit dan kompleks (Budiantara, 2011) Salah satu jenis regresi spline adalah regresi *smoothing splines* yang merupakan pemulusan data dengan menggunakan fungsi *splines* serta mempunyai jumlah kuadrat residual yang kecil. Estimasi fungsi *smoothing splines* diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS) sebagai berikut:

$$\sum_{i=1}^n (Y_i - f(xi))^2 + \lambda \int_0^1 [f''(x)]^2 dx \dots\dots\dots(7)$$

dengan  $\sum_{i=1}^n (Y_i - f(xi))^2$  merupakan jumlah kuadrat residual atau fungsi antara jarak data dengan dugaan sedangkan  $\int [f''(x)]^2$  merupakan ukuran kemulusan (*roughness penalty*), dan  $\lambda$  merupakan parameter pemulus yang bernilai diantara 0 dan 1. Parameter pemulus berfungsi sebagai pengontrol dimana jika  $\lambda$  semakin besar maka akan semakin kecil pula ragamnya.

Berbeda dengan regresi kernel yang mengoptimalkan *bandwidth* h agar menghasilkan kurva yang cocok dengan pola hubungan antara data, pada regresi *smoothing spline* yang perlu dioptimalkan yaitu *knot* dan parameter penghalus. Jumlah *knot* dan parameter penghalus yang optimal menurut (Eubank, 1999) dapat ditentukan dari nilai dari nilai GCV atau *Mean Squared Error* (MSE) yang minimum. Formula GCV dapat didefinisikan sebagai berikut:

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(\frac{1}{n}trace(I-H)\right)^2} \dots\dots\dots(8)$$

Dimana H merupakan  $X(X^T X)^{-1}X^T$  dan formula untuk MSE dapat didefinisikan sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots\dots\dots(9)$$

*Kuantil Regresi Smoothing Splines*

Regresi kuantil merupakan pengembangan dari regresi median (pada kuantil 0.5) untuk berbagai nilai kuantil. Regresi kuantil merupakan salah satu teknik regresi yang *robust* terhadap outlier (Koenker & Basset, 1978) Pada regresi kuantil hubungan antara variabel respon dan prediktor yang dijelaskan tidak hanya pada ukuran pemusatan (median bersyarat) variabel respon tetapi juga pada berbagai level kuantil yang dinotasikan  $\tau \in [0,1]$ . Menurut Andriyana (2015) regresi kuantil sangat baik digunakan pada sebaran data yang berdistribusi asimetris, padat pada ujung sebaran data, ataupun adanya outlier, karena estimator yang dihasilkan akan lebih efisien. Oleh karena itu, estimasi fungsi regresi kuantil *smoothing splines* diperoleh dengan meminimumkan fungsi sebagai berikut (Mulyani, 2017):

$$\sum_{i=1}^n \rho_\tau(Y_i - f(xi))^2 + \lambda \int_0^1 [f''(x)]^2 dx \dots\dots\dots(10)$$

dengan  $\rho_\tau = [\tau - 1(u < 0)]u$  yang merupakan *check function* dan  $f''(x)$  merupakan fungsi kontinu dalam selang  $[0,1]$  dengan bentuk fungsi sebagai berikut:

$$f(x) = \begin{cases} a_i + b_i(x - x_i) & x \in [x_i, x_{i+1}] \\ a_0 + b_0(x - x_0) & x \in [x_0, x_1] \end{cases} \dots\dots\dots (11)$$

dimana  $i = 1, 2, \dots, n$  sehingga  $a_i = y_i$  dan  $b_i = \frac{a_{i+1} - a_i}{x_{i+1} - x_i}$ . Jika nilai  $h_i = x_{i+1} - x_i$  dengan  $i = 1, 2, \dots, n - 1$  maka koefisien  $b_i = \frac{a_{i+1} - a_i}{h_i}$ .

Dengan melihat *constraint* maka dapat dikatakan bahwa  $a_0 = a_i$ ;  $b_0 = b_i$  dan  $b_{n-1} = b_n$ , maka persamaan fungsi  $f(x_{i+1}) = a_i + b_i h_i$  sehingga fungsi objektif regresi kuantil *smoothing splines* dengan meminimumkan:

$$\sum_{i=1}^n \rho_\tau(Y_i - f(x_i)) + \lambda \sum_{i=1}^{n-1} |f'(x_{i+1}) - f'(x_i)| \dots\dots\dots (12)$$

## Data dan Sumber Data

Data pada penelitian ini merupakan data sekunder yang bersumber dari *website* Badan Pusat Statistik (BPS) Provinsi Sumatera Utara (<https://sumut.bps.go.id>). Data yang digunakan antara lain data persentase penduduk miskin (tingkat kemiskinan) dan Indeks Pembangunan Manusia (IPM) tahun 2020 dengan unit analisis 33 yang kabupaten/kota yang berada di Provinsi Sumatera Utara. Adapun variabel respon pada penelitian ini adalah persentase penduduk miskin dan variabel prediktor adalah IPM.

## Tahapan Analisis

Tahapan analisis yang dilakukan pada penelitian adalah sebagai berikut:

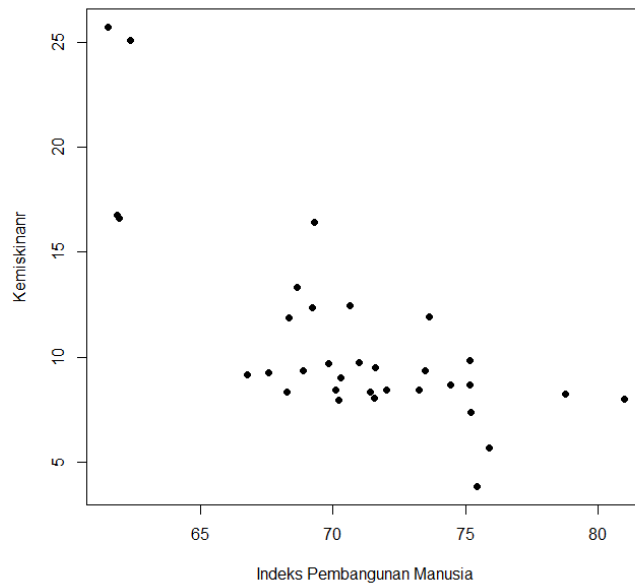
1. Melakukan identifikasi hubungan antara variabel respon dan variabel prediktor melalui *scatter plot* dan pengecekan *outlier* melalui *boxplot*.
2. Melakukan pemodelan menggunakan kernel dengan mencari *bandwidth* optimum dan pemodelan *smoothing splines* dengan mencari parameter pemulus yang optimum dilihat dari nilai GCV terkecil.
3. Melakukan pemodelan menggunakan regresi kuantil *smoothing splines*.
4. Membuat plot hasil regresi.
5. Melakukan perbandingan model.

## HASIL DAN PEMBAHASAN

### Identifikasi Model

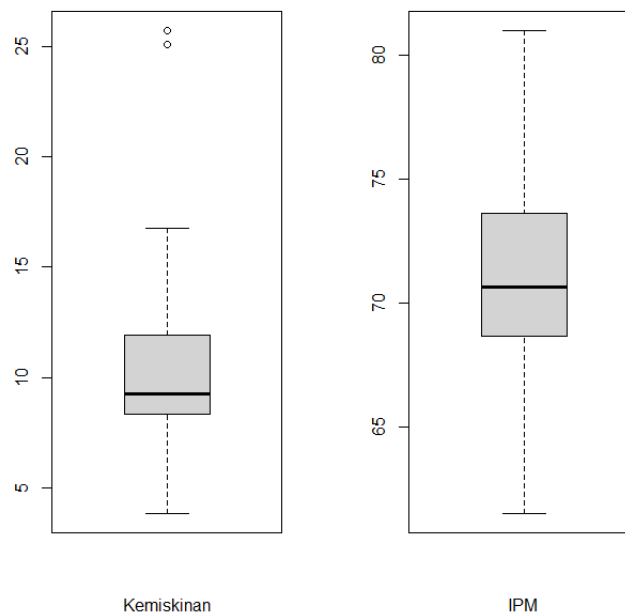
Sebelum melakukan identifikasi model, perlu diketahui gambaran data secara univariat. Berdasarkan data yang ada, dapat ditunjukkan bahwa rata-rata tingkat Kemiskinan di Provinsi Sumatera Utara pada tahun 2020 yaitu sebesar 10,8 persen dan standar deviasi sebesar 4,7 dengan nilai minimum 3,88 dan maksimum 25,69. Sementara rata-rata IPM di Provinsi Sumatera Utara pada tahun 2020 yaitu sebesar 70,75 dan standar deviasi 4,6 dengan nilai minimum 61,51 dan maksimum 80,98. Pada tahap awal dilakukan identifikasi model dari plot sebaran data menggunakan *scatter plot* Apabila hubungan antara variabel respon dan prediktor tidak membentuk pola tertentu atau tidak dapat ditentukan fungsi *slopenya*, maka pendekatan nonparametrik tepat untuk digunakan. Hasil *scatter plot* antara variabel kemiskinan dan IPM dapat dilihat pada Gambar 1.

Pada Gambar 1 dapat dilihat bahwa pola sebaran hubungan antara kemiskinan dan IPM tidak dapat diketahui fungsinya. Data cenderung mengumpul pada sisi bawah bagian tengah, namun terdapat juga beberapa sebaran di kiri atas dan kanan bawah. Oleh karena itu dirasa tepat untuk menggunakan regresi nonparametrik dalam melakukan pemodelan kemiskinan di Provinsi Sumatera Utara. Selanjutnya dilakukan identifikasi dengan melihat *boxplot* dari variabel-variabel penelitian yang disajikan pada Gambar 2. Dapat dilihat bahwa terdapat *outlier* pada variabel kemiskinan sehingga penggunaan regresi kuantil yang *robust* terhadap *outlier* juga tepat untuk dilakukan pada penelitian ini.



Sumber: hasil olah program R-4.0.2

Gambar 1. Scatter Plot Kemiskinan dan IPM



Sumber: hasil olah program R-4.0.2

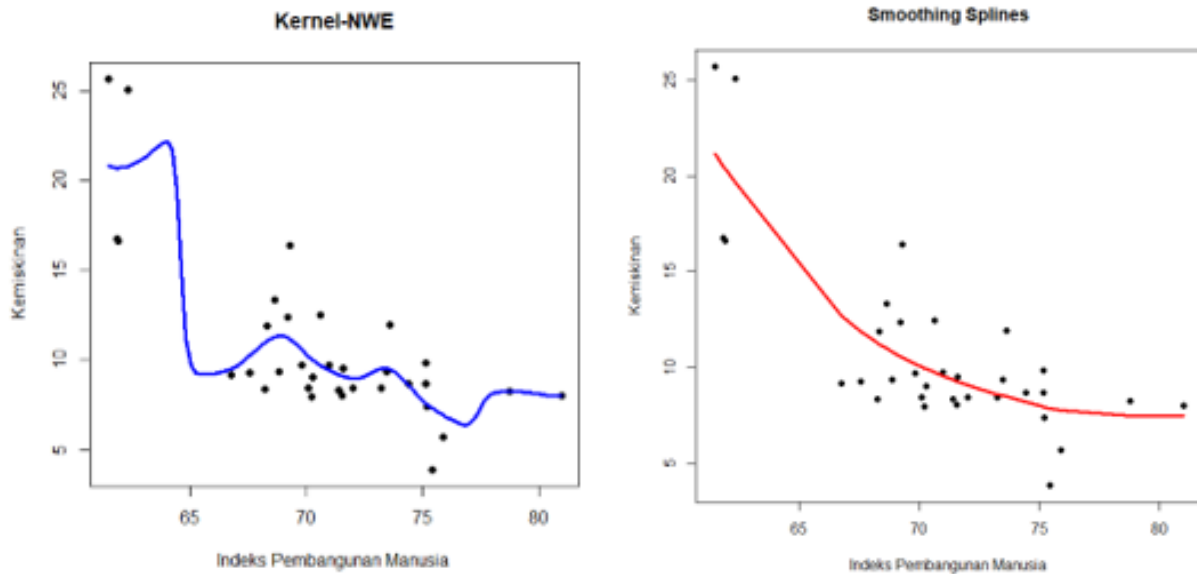
Gambar 2. Boxplot Variabel Kemiskinan dan IPM

### Pemodelan Kernel dan *Smoothing Splines*

Pemodelan regresi nonparametrik dengan kernel pada penelitian ini didasarkan pada fungsi densitas kernel *gaussian* (normal) dengan menggunakan estimator Nadaraya-Watson. Ketepatan dalam suatu pemulusan pada kernel bergantung pada dua hal, yaitu besarnya *bandwidth* dan fungsi kernel yang digunakan menjadi bobot.

Nilai *bandwidth* optimal yang diperoleh dari fungsi kernel gaussian dengan estimator Nadaraya-Watson adalah 2,13512 dengan GCV sebesar 11,78793. Sementara pada regresi nonparametrik *smoothing splines* yang kurvanya sangat bergantung dari nilai parameter pemulus ( $\lambda$ ). Semakin besar nilai parameter pemulus, maka akan semakin mulus bentuk kurvanya, namun hasil tersebut belum tentu lebih baik. Kurva yang baik, apabila kurva tersebut halus dan mengikuti pola hubungan datanya sehingga biasanya akan lebih kecil. Setelah dicobakan beberapa nilai parameter pemulus ( $\lambda$ ), maka nilai parameter pemulus yang optimal yaitu 0,00544

dengan GCV sebesar 47,29301. Kurva regresi nonparametrik kernel dan *smoothing splines* tersebut dapat dilihat pada Gambar 3.



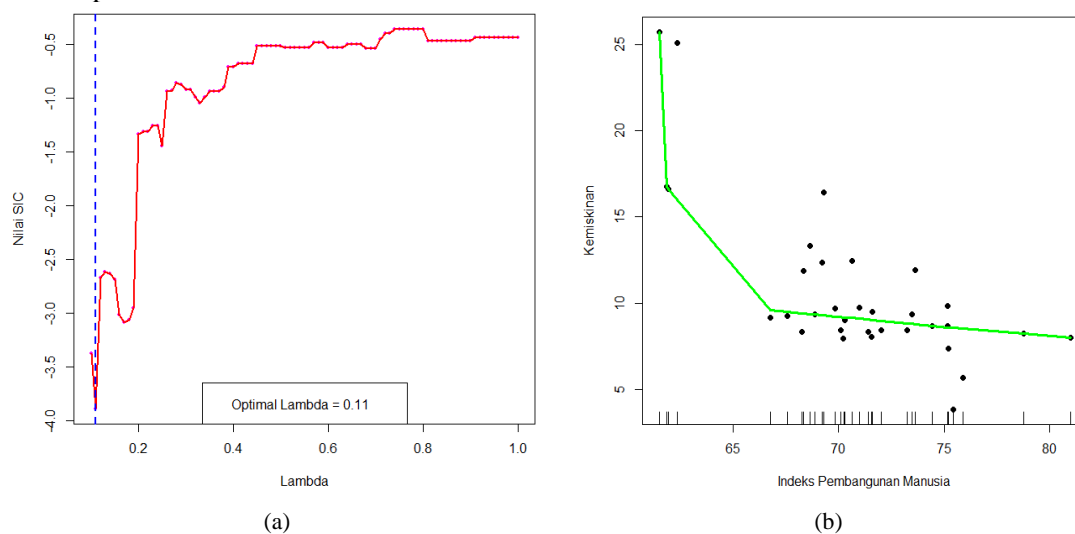
Sumber: hasil olah program R-4.0.2

Gambar 3. Kurva Regresi Kernel NWE (kiri) dan Smoothing Splines (kanan)

Berdasarkan Gambar 3 di atas menunjukkan bahwa kurva regresi kernel NWE lebih mengikuti dibandingkan dengan kurva *smoothing splines* tetapi kurva *smoothing splines* lebih halus dibanding dengan kurva kernel NWE. Hal ini masih saja terlihat belum sempurna dikarenakan terdapat *outlier* yang mempengaruhi pergerakan kurva. Oleh karena itu, dalam penelitian ini dilanjutkan dengan menggunakan regresi kuantil *smoothing splines*.

### Pemodelan Regresi Kuantil *Smoothing Splines*

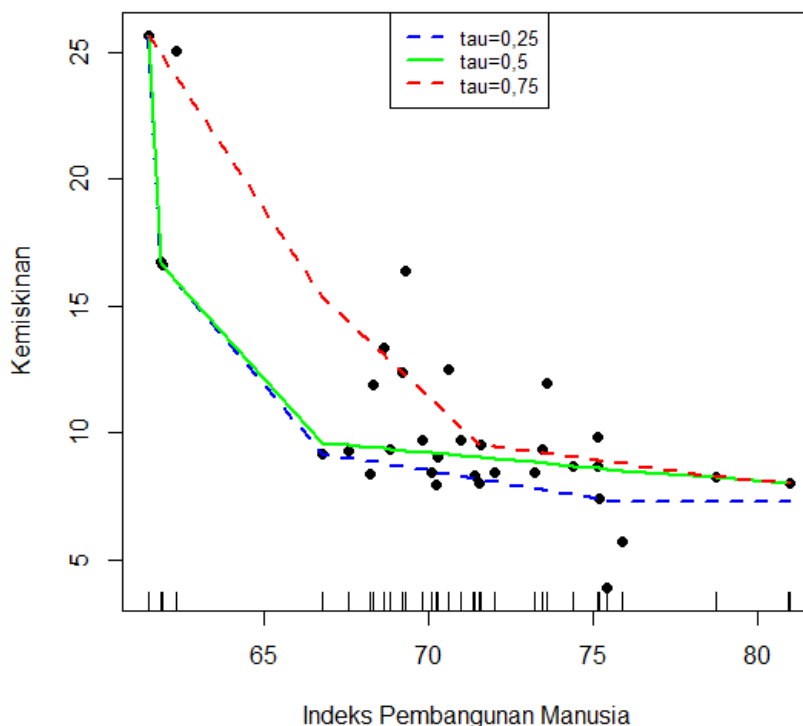
Pada pemodelan kernel dan *smoothing splines* sebelumnya diketahui bahwa terdapat beberapa kekurangan dari kurva regresi kernel dan *smoothing splines*. Kekurangan dari kurva regresi kernel yaitu kurang halus (under smooth) sementara kekurangan dari kurva *smoothing splines* yaitu kurva tidak dapat mengakomodir keberadaan outlier sehingga dilakukan pemodelan dengan regresi kuantil *smoothing splines*. Dalam regresi kuantil *smoothing splines* perlu dilakukan kembali penentuan parameter penghalus yang optimum, yaitu menggunakan nilai kriteria *Schward Information Criteria* (SIC) dan nilai  $\tau = 0$ , (Meimela, 2020). Hasil penentuan parameter penghalus dan pemodelan regresi kuantil *smoothing splines* nilai  $\tau = 0,5$  dapat dilihat pada Gambar 4.



Sumber: hasil olah program R-4.0.2

Gambar 4. a) Plot SIC dan Parameter Pemulus, b) Kurva Regresi Kuantil *Smoothing Splines*

Berdasarkan Gambar 4a dapat diketahui bahwa parameter pemulus yang optimal pada regresi kuantil *smoothing splines* adalah 0,11. Kemudian dengan nilai parameter pemulus tersebut dilakukan pemodelan menggunakan regresi kuantil *smoothing splines* dan hasilnya terlihat pada gambar 4b, dapat dilihat bahwa kurva tidak *under smooth* atau *over smooth* dan juga dapat mengakomodir keberadaan *outlier*. Oleh karena itu, dapat dilihat kurva regresi menggunakan regresi kuantil *smoothing splines* terlihat lebih mengikuti sebaran data. Regresi kuantil *smoothing splines* seperti dijelaskan sebelumnya dapat menampilkan beberapa level kuantil, kurva regresi kuantil *smoothing splines* dengan tiga kuantil nilai  $\tau = (0,25, 0,5, \text{ dan } 0,75)$  dapat dilihat pada Gambar 5.



Sumber: hasil olah program R-4.0.2

Gambar 5. Kurva Regresi Kuantil Smoothing Splines dengan Tiga Kuantil  $\tau = (0,25, 0,5, \text{ dan } 0,75)$

Berdasarkan Gambar 5 dapat dilihat bahwa regresi kuantil *smoothing splines* dapat memberikan kurva regresi yang berbeda-beda, tergantung dari kuantil ke-berapa yang ingin dilihat. Hal ini juga merupakan kelebihan dari regresi kuantil. Nilai koefisien determinasi dari model regresi kuantil *smoothing splines*  $R^2$  sebesar 67,63 persen, yang berarti variasi dari variabel kemiskinan yang dapat dijelaskan oleh IPM adalah sebesar 67,63 persen dan sisanya 32,37 persen dijelaskan oleh variabel lain diluar model.

### Perbandingan Model

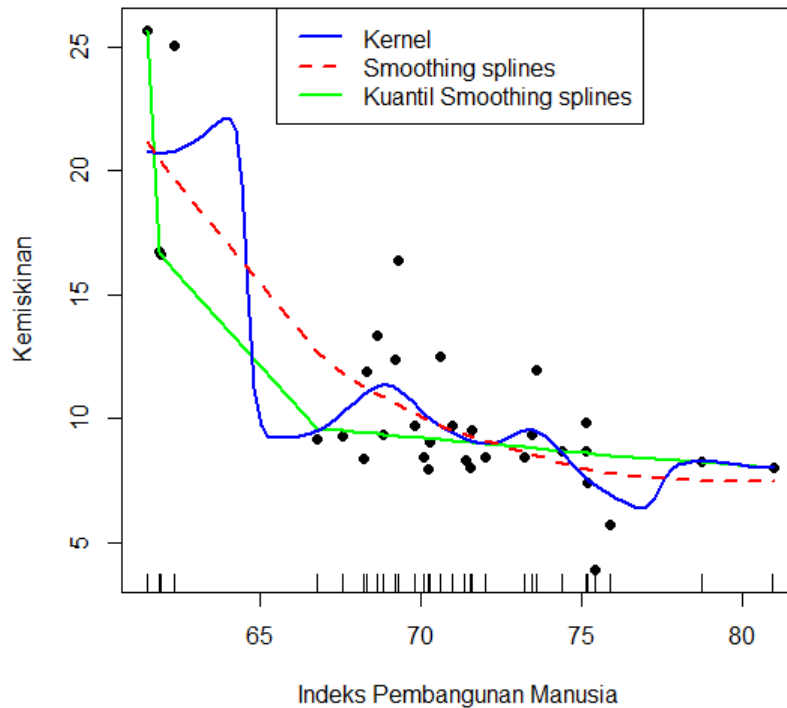
Untuk melakukan perbandingan model dilakukan perbandingan nilai GCV dan *Root Mean Squared Error* (RMSE), nilai GCV dan RMSE dapat dilihat pada Tabel 1 berikut:

Tabel 1. Perbandingan Model Regresi Nonparametrik

No.	Metode	GCV	RMSE
1.	Kernel NWE	11,78793	3,329314
2.	<i>Smoothing Splines</i>	47,29301	6,668598
3.	Kuantil <i>Smoothing Splines</i>	3,81497	1,894007

Sumber: hasil olah program R-4.0.2

Berdasarkan Tabel 1 nilai GCV dan RMSE terkecil yaitu pada model regresi kuantil *smoothing splines* yang menandakan bahwa model ini merupakan model terbaik dibandingkan dengan kernel NWE dan *smoothing splines*. Hal ini disebabkan metode kuantil *smoothing splines* yang bersifat *robust* terhadap outlier. Jika dibandingkan antara metode Kernel NWE dibandingkan dengan *smoothing splines* maka model kernel NWE lebih baik jika dilihat dari nilai GCV dan RMSEnya. Selain itu, perbandingan antara ketiga model juga dapat dilihat melalui kurva regresinya yang terdapat pada Gambar 6.



Sumber: hasil olah program R-4.0.2

Gambar 6. Perbandingan Kurva Regresi Nonparametrik

Pada Gambar 6 dapat dilihat perbedaan kurva regresi dari ketiga metode yang digunakan dalam penelitian ini. Kurva *smoothing splines* yang berwarna merah merupakan kurva regresi yang paling halus (*smooth*) dibanding dengan yang lainnya, namun tidak terlalu mengikuti pola sebaran hubungan data, khususnya pada data yang *outlier* sementara kurva kernel NWE merupakan kurva yang kurang halus dan kurva kuantil *smoothing splines* yang paling mengikuti sebaran dari hubungan data. Oleh karena itu dapat disimpulkan bahwa kurva kuantil *smoothing splines* merupakan metode yang terbaik dibandingkan dengan yang lainnya.

## KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan dapat disimpulkan bahwa:

1. Pemodelan kemiskinan dapat dilakukan dengan analisis regresi nonparametrik dikarenakan sebaran hubungan data sulit untuk diidentifikasi.
2. Pemodelan menggunakan regresi nonparametrik kernel diperoleh *bandwidth* optimal dari fungsi kernel *gaussian* dengan estimator Nadaraya-Watson adalah 2,13512 dengan GCV sebesar 11,78793, pemodelan menggunakan regresi nonparametrik *smoothing splines* diperoleh nilai parameter pemulus yang optimal yaitu 0,00544 dengan GCV sebesar 47,29301, dan pemodelan dengan regresi kuantil *smoothing splines* diperoleh parameter pemulus yang optimal yaitu 0,11 dengan GCV sebesar 3,81497.
3. Hasil perbandingan model diperoleh bahwa metode regresi kuantil *smoothing splines* merupakan metode terbaik dalam melakukan pemodelan kemiskinan di Provinsi Sumatera Utara karena memiliki kurva regresi yang lebih mengikuti sebaran hubungan data dan nilai GCV dan RMSE yang lebih rendah dibandingkan metode lainnya. Besarnya variasi dari variabel kemiskinan yang dapat dijelaskan oleh IPM dalam model regresi kuantil *smoothing splines* adalah sebesar 67,63 persen dan sisanya 32,37 persen dijelaskan oleh variabel lain diluar model

## DAFTAR PUSTAKA

- Andriyana, Y. (2015). P-Splines Quantile Regression in Varying Coefficient Models. Disertasi KU Leuven. Belgia.
- Badan Pusat Statistik (BPS) diakses dari <http://www.sumut.bps.go.id/>, diakses pada tanggal 30 Juli 2021 pada jam 11.03 WIB.



- Badan Pusat Statistik. (2021a). Berita Resmi Statistik Profil Kemiskinan di Indonesia Maret 2021. Badan Pusat Statistik. Jakarta: No. 53/07/Th. XXIV, 15 Juli 2021.
- Badan Pusat Statistik. (2021b). Indeks Pembangunan Manusia 2020. Jakarta: Badan Pusat Statistik.
- Budiantara, I. N. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter. Seminar Nasional FMIPA Undiksha.
- Dowling, J. M., & Valenzuela, M. R. (2010). *Economic Development in Asia Second Edition*. Singapore: CENGAGE Learning.
- Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- Hardle. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
- Hoelman, M. B., Parhusip, B. P., Eko, S., Bahagijo, S., & Santono, H. (2015). *Panduan SDGs Untuk Pemerintah Daerah (Kota dan Kabupaten) dan Pemangku Kepentingan Daerah*. Jakarta: Infid.
- Koenker, R., & Basset, J. (1978). Regression Quantile. *Econometrica* 46, 1, 33-50.
- Meimela, A. (2020). Model Hubungan Jumlah Pengangguran dan Indeks Kedalaman Kemiskinan di Pulau Sumatera Tahun 2019 Menggunakan Regresi Nonparametrik Splines. *Jurnal Ilmu Ekonomi dan Pembangunan (JIEP)*, Vol 20, No 2. doi:<https://doi.org/10.20961/jiep.v20i2.41701>.
- Mulyani, S. (2017). *Pemodelan Hubungan Indeks Pembangunan Manusia dan Persentase Penduduk Miskin Menggunakan Regresi Kuantil Smoothing Splines*. Tesis Universitas Padjadjaran. Bandung. 93 hlm.
- Prasetyoningrum, A. K., & Sukmawati, U. (2018). Analisis Pengaruh Indeks Pembangunan Manusia (IPM), Pertumbuhan Ekonomi dan Pengangguran Terhadap Kemiskinan di Indonesia. *EQUILIBRIUM: Jurnal Ekonomi Syariah*, Volume 6, Nomor 2, 217 - 240.
- Safuridar, & Putri, N. I. (2019). Pengaruh Indeks Pembangunan Manusia, Pengangguran Dan Jumlah Penduduk Terhadap Tingkat Kemiskinan di Aceh Bagian Timur. *Jurnal Samudra Ekonomika*, Volume 3, Nomor 1
- Sihite, L. (2020). *Pengaruh Dana Desa, IPM, dan Pertumbuhan Ekonomi Terhadap Penurunan Tingkat Kemiskinan di Provinsi Sumatera Utara*. Tesis Universitas Sumatera Utara. Medan. 112 hlm.
- Todaro, M. P. (2006). *Pembangunan Ekonomi (edisi kesembilan, jilid 1)*. Jakarta: Erlangga.