

MODEL REGRESI NONPARAMETRIK BERDASARKAN ESTIMATOR POLINOMIAL LOKAL KERNEL PADA KASUS PERTUMBUHAN BALITA

¹Mifta Luthfin Alfiani, ²Indah Manfaati Nur, ³Tiani Wahyu Utami

^{1,2,3}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang
Alamat e-mail : Pinky7pipin@gmail.com

ABSTRAK

Analisis regresi merupakan salah satu tehnik analisis data dalam statistika yang paling banyak digunakan untuk mengetahui hubungan antara variabel prediktor (X) dengan variabel responnya (Y). Dalam analisis regresi, kurva regresi dapat didekati dengan model regresi nonparametrik. Model regresi nonparametrik yang pengamatan dilakukan sebanyak n subjek yang saling independen pada waktu satu titik tertentu sehingga mengikuti struktur data *cross-section*. Dalam penelitian ini akan dikaji mengenai model regresi nonparametrik berdasarkan estimator polinomial lokal kernel pada pertumbuhan balita. Estimator polinomial lokal kernel dapat diperoleh dengan meminimumkan WLS (*Weighted Least Square*). Estimator polinomial lokal kernel sangat bergantung pada *bandwidth* (h) optimal. Penentuan *bandwidth* optimal dapat diperoleh dengan menggunakan metode GCV (*Generalized Cross Validation*). Selanjutnya diterapkan pada aplikasi model regresi nonparametrik berdasarkan estimator polinomial lokal kernel pada kasus pertumbuhan balita untuk menganalisis pengaruh antara berat badan dan umur. Pemodelan dilakukan pada balita laki-laki dan perempuan, pada balita laki-laki mempunyai nilai MSE = 1,487844 dan koefisien determinasi (R^2) = 81,86%, pada balita perempuan mempunyai nilai MSE = 1,388796 dan koefisien determinasi (R^2) = 79,55%.

Kata Kunci : Regresi Nonparametrik, Estimator Polinomial Lokal Kernel, WLS, GCV, Pertumbuhan Balita.

PENDAHULUAN

Metode regresi merupakan metode statistika untuk mengetahui hubungan antara variabel respon dengan variabel prediktor. $f(x_i)$ merupakan kurva regresi yang dapat didekati dengan tiga pendekatan yaitu pendekatan regresi parametrik, pendekatan regresi nonparametrik, dan pendekatan regresi semiparametrik^[5].

Pada pendekatan regresi parametrik diasumsikan bahwa bentuk kurva regresi diketahui berdasarkan informasi sebelumnya dan teori, ataupun pengalaman masa lalu. Akibatnya estimator kurva

regresi diperoleh dengan mengestimasi parameternya^{[6];[2]}. Pendekatan regresi nonparametrik tidak memberikan asumsi bentuk kurva tertentu ataupun tidak ada informasi mengenai bentuk kurva regresi. Kurva regresi dapat diasumsikan mulus atau *smooth*, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektifitas peneliti^{[5];[9]}. Jika variabel respon diketahui hubungannya dengan salah satu variabel prediktor, tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Dalam kondisi seperti

ini^{[10];[9]} menyarankan menggunakan regresi semiparametrik. Regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik^{[4];[9]}.

Jika n pengamatan independen $\{X_i, Y_i\}_{i=1}^n$ $i=1,2,3,\dots,n$. Hubungan antara X_i dan Y_i mengikuti model regresi nonparametrik. Bentuk kurva regresi tidak diketahui di asumsikan *smooth*. Estimator kurva regresi diperoleh dengan mengestimasi parameternya. Terdapat beberapa pendekatan untuk mengestimasi kurva regresi salah satunya dengan estimator polinomial lokal kernel. Kelebihan estimator polinomial lokal kernel adalah dapat mengurangi asimtotik bias dan menghasilkan estimasi yang baik^[11]. Estimator polinomial lokal kernel dapat diperoleh dengan optimasi WLS (Weighted Least Square). Sedangkan untuk mengestimasi parameter penghalus (*bandwidth*) menggunakan metode GCV (*Generalized Cross Validation*).

Penelitian sebelumnya telah banyak dikembangkan mengenai estimasi model regresi nonparametrik pada data longitudinal berdasarkan estimator polinomial lokal kernel GEE^[8], pendekatan regresi semiparametrik untuk data longitudinal terhadap kadar trombosit demam berdarah dengue^[9].

Peningkatan secara bertahap dari tubuh, organ, dan jaringan masa konsepsi sampai remaja dengan pengertian usia anak dibawah lima tahun^[7]. Masa balita merupakan periode terpenting dari dalam proses tumbuh kembang manusia. Mengingat masa tumbuh kembang di usia balita yang berlangsung cepat dan tidak akan pernah terulang lagi. Apabila anak usia balita mendapatkan stimulus yang baik, maka seluruh aspek pertumbuhan anak balita akan tumbuh secara optimal. Cara mudah untuk mengetahui baik tidaknya pertumbuhan balita adalah dengan mengamati grafik pertumbuhan berat badan yang terdapat pada Kartu Menuju Sehat (KMS). KMS digunakan

untuk memantau keadaan kesehatan dan gizi melalui pertumbuhan atas dasar kenaikan berat badan. Kartu ini merupakan gambar kurva berat badan anak usia 0-5 tahun terhadap umurnya. Berat badan menurut umur mempunyai pengaruh terhadap perubahan keadaan gizi^{[1];[3]}. Penggunaan variabel berat badan menurut umur akan mudah dipahami oleh masyarakat umum terhadap pertumbuhan balita.

Pada tulisan ini akan dibahas bagaimana mengestimasi kurva regresi nonparametrik menggunakan estimator polinomial lokal kernel. Selanjutnya menerapkan model regresi tersebut pada data pertumbuhan balita berdasarkan berat badan menurut umur.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Posyandu Desa Truwolu, Grobogan, Jawa Tengah. Data tersebut mengenai pertumbuhan berat badan (y) yang dipengaruhi umur balita (x).

Metode Analisis

Langkah-langkah yang dilakukan pada penelitian ini adalah :

1. Mengestimasi model regresi nonparametrik berdasarkan estimator polinomial lokal kernel dengan langkah-langkah sebagai berikut :

- a. Diberikan data observasi (y_i, x_i) , yang memenuhi regresi nonparametrik.

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, 3, \dots, n \quad (1)$$

- b. Menyatakan (y_i) dapat didekati oleh polinomial berderajat p

$$m(x_i) \approx m(x_i) + (x_i - x)m^{(1)} + \frac{(x_i - x)^2 m^{(2)}}{2!} + \dots + \frac{(x_i - x)^p m^{(p)}(x)}{p!}$$

$$\text{Jika } \beta_r(x) = \frac{m^{(r)}(x)}{r!} \text{ maka}$$

$$m(x_i) = \beta_0(x) + (x_i - x)\beta_1(x_1) + \dots + (x_i - x)^p \beta_p(x)$$

- c. Dengan α mengestimasi β dengan optimasi atau meminimumkan kriteria *Weighted Lest Square* :

$$M = \sum_{i=1}^n (y_i - m(x_i))^2 K_h(x_i - x)$$

Dengan $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ fungsi pembobot Kernel

- d. Menyatakan $\hat{m} \approx \mathbf{X}\hat{\beta}$
 Dengan \mathbf{X} adalah matrik berukuran $1 \times (p+1)$ dan \hat{m} adalah vektor berukuran 1×1 .
 e. Mendapatkan bentuk matrik $\mathbf{A}(h)$ berukuran $N \times N$ dengan cara menyelesaikan persamaan berikut :
 $\hat{m} = \mathbf{A}(h)y^*$

2. Hasil estimasi yang diperoleh diterapkan untuk menganalisis pengaruh variabel umur terhadap variabel berat badan (BB) di Posyandu Desa Truwolu, Grobogan Jawa Tengah dengan langkah-langkah sebagai berikut :

- a. Membuat plot data berpasangan (y_i, x_i) , $i = 1, 2, \dots, n$
 b. Menentukan jenis pembobot dan fungsi Kernel Gaussian.
 c. Menentukan matriks $\mathbf{A}(h)$ berukuran $N \times N$
 d. Memilih orde polinomial p dan nilai *bandwidth* optimal yang meminimumkan

$$GCV = \frac{n^{-1} \sum_{i=1}^n [y_i - \hat{y}_i]^2}{(n^{-1} \text{tr} [I - \mathbf{A}(h)])^2}$$

- e. Memodelkan orde polinomial lokal p dan nilai *bandwidth* optimal dari langkah c secara simultan.
 f. Menghitung nilai MSE
 $MSE(h) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)$
 g. Mendapatkan model pertumbuhan berat badan terhadap umur dengan polinomial lokal kernel.

HASIL PENELITIAN

Berdasarkan tujuan penelitian, maka pada bab ini akan dilakukan estimasi model regresi nonparametrik berdasarkan estimator polinomial lokal kernel dengan meminimumkan *Weighted Least Square* (WLS). Hasil estimasi yang diperoleh pada data pertumbuhan berat badan terhadap umur di Posyandu desa Truwolu, Grobogan, Jawa Tengah.

Estimasi Model Regresi Nonparametrik Berdasarkan Estimator polinomial Lokal Kernel.

Diberikan data observasi (y_i, x_i) yang memenuhi regresi nonparametrik, dengan y_i sebagai variabel respon dan x_i sebagai variabel prediktor. Hubungan y_i dengan x_i tidak diketahui bentuk fungsinya. Maka kita dapat memodelkan hubungan antara y_i dan x_i dapat dimodelkan dengan menggunakan model regresi nonparametrik sebagai berikut :

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, 3, \dots, n$$

Model (1) dapat ditulis dalam bentuk matriks :

$$\mathbf{y} = \mathbf{m} + \boldsymbol{\varepsilon} \tag{2}$$

Dengan

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ dan}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Fungsi $m(x_i)$ tidak diketahui bentuk fungsinya yang disebut dengan fungsi regresi. Misalkan X adalah variabel prediktor dimana fungsi $m(x_i)$ akan diestimasi dengan estimator polinomial lokal. Dengan deret Taylor, $m(x_i)$ dapat didekati oleh polinomial berderajat p sebagai berikut :

$$m(x_i) = \beta_0(x) + (x_i - x)\beta_1(x_1) + \dots + (x_i - x)^p \beta_p(x) \tag{3}$$

Persamaan (2) dapat ditulis menjadi :

$$m = X\beta$$

Dengan

$$X = \begin{bmatrix} x(h_1) & 0 & \dots & 0 \\ 0 & x(h_2) & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & x(h_n) \end{bmatrix},$$

$$X(h_i) = \begin{bmatrix} 1 & (x_1 - x) & \dots & (x_1 - x)^p \\ 1 & (x_2 - x) & \dots & (x_2 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x) & \dots & (x_n - x)^p \end{bmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \beta_i = [\beta_0, \beta_1, \dots, \beta_p]^T$$

$\hat{\beta}$ dapat diperoleh dengan cara meminimumkan *Weighted Least Square* (WLS)

$$S(\beta) = (y - X\beta)^T K_h (y - X\beta) \quad (4)$$

$$\text{Dengan } y_i = [y_1, y_2, \dots, y_n]$$

$$X = [x_1^T, \dots, x_n^T]$$

$$K_h = \text{diag}(K_{h1}, K_{h2}, \dots, K_{hn}),$$

$$K_{hi} = \text{diag}(K_{hi}(x_1 - x), K_{hi}(x_2 - x), \dots, K_{hi}(x_n - x))$$

Matriks K_h adalah matriks yang berisi pembobot, dengan $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$. $K(\cdot)$ adalah fungsi Kernel, dan h adalah *bandwidth*.

Nilai estimasi β yaitu $\hat{\beta}$ yang bila disubstitusikan kedalam persamaan (4) akan meminimumkan $S(\beta)$ diperoleh :

$$\hat{\beta} = (X^T K_h X)^{-1} X^T K_h y \quad (5)$$

Dengan demikian m dapat dinyatakan sebagai berikut :

$$\hat{m} = X\hat{\beta}$$

$$= X(X^T K_h X)^{-1} X^T K_h y$$

Pemodelan Pertumbuhan Balita Dengan Menggunakan Estimator Polinomial Lokal Kernel

Estimasi model regresi nonparametrik dengan menggunakan estimator polinomial lokal kernel untuk data pertumbuhan balita di Posyandu Desa Truwolu, Grobogan, Jawa Tengah. Pada penelitian ini, difokuskan untuk mengetahui hubungan antara berat badan terhadap umur.

Estimasi model regresi nonparametrik polinomial lokal kernel sangat bergantung pada pemilihan *bandwidth* optimal dan derajat polinomial, dimana penentuan *bandwidth* optimal dan derajat polinomial dapat dilihat pada nilai GCV minimum. Pada proses estimasi ini, *bandwidth* optimal dan nilai GCV minimum dihitung pada orde polinomial satu sampai empat ($p= 1, p= 2, p= 3, p= 4$). Pemilihan *bandwidth* optimal dan orde polinomial yang digunakan dalam proses estimasi dilakukan dengan memilih nilai GCV minimum pada masing-masing orde polinomial dan *bandwidth*. Dalam penelitian ini fungsi Kernel yang digunakan adalah fungsi Kernel Gaussian.

Pemilihan Nilai *Bandwidth* dan Orde Polinomial pada Model Regresi Nonparametrik

Langkah pertama sebelum mengestimasi model regresi secara simultan adalah menentukan *bandwidth* optimal dan orde polinomial. Metode GCV merupakan metode yang digunakan dalam menentukan pemilihan *bandwidth* optimal dan orde polinomial. Langkah awal dalam pemilihan *bandwidth* optimal dan orde polinomial adalah menentukan nilai awal *bandwidth* dan orde polinomial. Nilai awal dari penentuan *bandwidth* optimal dan orde polinomial digunakan sebagai acuan dalam estimasi model regresi nonparametrik pada balita laki-laki maupun perempuan.

Pada balita laki-laki orde polinomial dengan nilai GCV minimum adalah 1.632917, dengan nilai *bandwidth* (h) sebesar 53.3. Dengan demikian, model

regresi nonparametrik untuk balita laki-laki didekati dengan Polinomial Lokal dengan nilai awal *bandwidth* sebesar 53.3 dan orde polinomial $p=4$.

Pada balita perempuan orde polinomial dengan nilai GCV minimum adalah 1.524211, dengan nilai *bandwidth* (h) sebesar 57.9. Dengan demikian, model regresi nonparametrik untuk balita perempuan didekati dengan Polinomial Lokal dengan nilai awal *bandwidth* sebesar 57.9 dan orde polinomial $p=4$.

Model Regresi Nonparametrik Berdasarkan Estimator Polinomial Lokal Kernel Pada Balita Laki-laki dan Balita Perempuan

Pemodelan pada balita laki-laki dan balita perempuan akan digunakan fungsi Kernel Gaussian, sehingga diperoleh model sebagai berikut :

Pada balita laki-laki, estimasi fungsi regresi nonparametrik berdasarkan estimator polinomial lokal kernel diperoleh MSE sebesar 1.487844 dan nilai koefisien determinasi (*R-square*) sebesar 0.8186386.

Pada balita perempuan, estimasi fungsi regresi nonparametrik berdasarkan estimator polinomial lokal kernel diperoleh MSE sebesar 1.388796 dan nilai koefisien determinasi (*R-square*) sebesar 0.7955247.

KESIMPULAN

Berdasarkan hasil penelitian, maka dapat disimpulkan bahwa :

1. Estimasi model regresi nonparametrik dengan menggunakan estimator polinomial lokal kernel pada kasus pertumbuhan balita adalah sebagai berikut :

$$y = m + \varepsilon$$

dengan $\hat{m} = X\hat{\beta}$

$$\hat{\beta} = (X^T K_h X)^{-1} X^T K_y y$$

2. Dari hasil penerapan model regresi nonparametrik yaitu pada data pertumbuhan balita di desa truwolu, Grobogan, Jawa Tengah, diperoleh dari hasil pemodelan pada balita laki-laki nilai *bandwidth* optimal sebesar 53,3 dengan orde polinomial $p=4$, MSE sebesar 1,487844 dan koefisien determinasi sebesar 81,86%, sedangkan pemodelan pada balita perempuan diperoleh *bandwidth* optimal sebesar 57,9 dengan orde polinomial $p=4$, MSE sebesar 1,388798 dan koefisien determinasi sebesar 79,55%, dan. *Scatterplot* untuk balita laki-laki maupun perempuan tampak bahwa pada selang usia awal bulan sampai akhir bulan mengindikasikan kurva mempunyai bentuk pola tertentu dan plot data menunjukkan adanya perubahan perilaku kurva pada selang waktu tertentu.

Dalam penelitian ini, permasalahan yang dikaji masih sangat terbatas hanya oleh karena itu perlu penelitian lebih lanjut yaitu pada estimasi model pertumbuhan balita berat badan menurut umur dapat digunakan faktor lain yang mempengaruhi pertumbuhan balita, misalnya berat badan orang tua, masukkan gizi, kondisi sosial serta ekonomi, dan riwayat penyakit.

UCAPAN TERIMA KASIH

Pada penelitian ini saya ucapkan terima kasih kepada Posyandu di desa Truwolu, Grobogan, Jawa Tengah atas data yang telah diberikan.

DAFTAR PUSTAKA

[1] Aritonang, I. 2000. *Pemantauan Pertumbuhan Balita (Petunjuk Praktis Menilai Status Gizi & Kesehatan)*, Kanisius, Yogyakarta.

-
-
- [2] Budiantara, I.N. Lestari, B., dan Islamiyati, A., 2010. *Estimator Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Heterokedastisitas untuk Data Longitudinal*, Hibah Penelitian Kompetensi, DP2M-Dikti: Jakarta.
- [3] Dedy, S. 2005. *Pendekatan Kernel untuk Estimasi Kurva Regresi Nonparametrik Studi Kasus Pertumbuhan Anak Balita*. Skripsi. Jurusan Statistika.FMIPA ITS. Surabaya.
- [4] Engle, R.F., Granger, C.W. J., Rice, J., dan Weiss, A., 1986. *Semiparametric Estimates of The Relation Between Weather and Electricity Sales*. Journal of the American Statistical Association. Vol.81, hal 310-320.
- [5] Eubank, R.M., 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker. New York.
- [6] Hardle, W., 1990. *Applied Nonparametric Regression*. Cambridge University Press. New York.
- [7] Jelliffe. 1989. *Community Nutritional Assesment with Special Reference to Less Technically Developed Countries*. Oxford.
- [8] Utami, T.W. 2010. *Estimasi Model Regresi Nonparametrik Pada Data Longitudinal Berdasarkan Estimator Polinomial Lokal Kernel Generalized Estimating Equation*. Skripsi. Surabaya. Departemen Matematika Fakultas Sains dan Teknologi Universitas Airlangga.
- [9] Utami, T.W. 2013. *Pendekatan Regresi Semiparametrik Polinomial Lokal Untuk Data Longitudinal Terhadap Kadar Trombosit Demam Berdarah Dengue*. Tesis. Surabaya. Program Magister Jurusan Statistika FMIPA Institut Teknologi Sepuluh November.
- [10] Wahba, G. 1990. *Spline Models for Observation Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics. Vol. 59.
- [11] Welsh, A.H dan Yee, T.Y. 2005. *Local Regression for Vector Responses*. *Journal of Statistical Planning and Inference*.136 (2006) 3007-3031.