ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR NADARAYA-WATSON DENGAN FUNGSI KERNEL *TRIANGLE*

(Skripsi)

Oleh

AULIA PUTRI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

ABSTRACT

ESTIMATION OF NONPARAMETRIC REGRESSION MODEL USING NADARAYA-WATSON ESTIMATOR WITH KERNEL TRIANGLE FUNCTION

By

AULIA PUTRI

Nonparametric regression is one of the statistical methods that is used to determine the pattern of the relationship between predictor variables with response variables whose form of function is unknown. The nonparametric regression curve is only assumed to be smooth while the data will look for its own estimation form. The kernel function is one approach to estimate the nonparametric regression curve. This study aims to estimate the nonparametric regression curve using the Nadaraya-Watson estimator with the kernel Triangle function on toddler growth data in Sidorejo Village, Kec. Sidomulyo South Lampung and simulation data. The results shows that the optimal regression curve estimation for real data is having bandwidth 5,0 with GCV 1,363146 while in the simulation data obtained the optimal bandwidth 5,0 with GCV 1,867415.

Keywords: Kernel Nonparametric Regression, Bandwidth, GCV

ABSTRAK

ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR NADARAYA-WATSON DENGAN FUNGSI KERNEL *TRIANGLE*

Oleh

AULIA PUTRI

Regresi nonparametrik adalah salah satu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon yang tidak diketahui bentuk fungsinya. Kurva regresi nonparametrik hanya diasumsikan mulus dimana data akan mencari bentuk estimasinya sendiri. Fungsi kernel merupakan salah satu pendekatan untuk mengestimasi kurva regresi nonparametrik. Penelitian ini bertujuan untuk mengestimasi kurva regresi nonparametrik menggunakan estimator Nadaraya-Watson dengan fungsi kernel *Triangle* pada data pertumbuhan balita di Desa Sidorejo Kec. Sidomulyo Lampung Selatan dan data simulasi. Hasil penelitian menunjukkan estimasi kurva regresi yang optimal memiliki *bandwidth* 5,0 dengan GCV 1,363146 sedangkan pada data simulasi diperoleh *bandwidth* yang optimal 5,0 dengan GCV 1,867415.

Kata Kunci: Regresi Nonparametrik Kernel, *Bandwidth*, GCV

ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR NADARAYA-WATSON DENGAN FUNGSI KERNEL *TRIANGLE*

Oleh

AULIA PUTRI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

Judul Skripsi : **ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR NADARAYA-WATSON DENGAN FUNGSI KERNEL *TRIANGLE***

Nama Mahasiswa : *Aulia Putri*

Nomor Pokok Mahasiswa : 1517031055

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Dr. Aang Nuryaman, S.Si., M.Si.**
NIP. 19650125 199003 2 001	NIP. 19740316 200501 1 001

2. Mengetahui

Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Sekretaris : **Dr. Aang Nuryaman, S.Si.,M.Si.** ....

Penguji
Bukan Pembimbing : **Dr. Khoirin Nisa, S.Si., M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
UNIVERSITAS LAMPUNG
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

Drs. Suratman, M.Sc.
NIP. 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi: **06 Agustus 2019**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan dibawah ini:

Nama : **Aulia Putri**

Nomor Pokok Mahasiswa : **1517031055**

Jurusan : **Matematika**

Judul : **ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR NADARAYA-WATSON DENGAN FUNGSI KERNEL *TRIANGLE***

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain, dan semua hasil tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 06 Agustus 2019

Penulis

Aulia Putri
NPM. 1517031055

RIWAYAT HIDUP

Penulis bernama lengkap Aulia Putri, anak pertama dari dua bersaudara yang dilahirkan di Kalianda pada tanggal 12 Oktober 1997 oleh pasangan Bapak Sobirin dan Ibu Iswanti.

Penulis menyelesaikan pendidikan taman kanak-kanak di TK Al-Khoiriyah Sidomulyo pada tahun 2003. Sekolah dasar di SD Negeri 05 Sidorejo pada tahun 2009. Sekolah menengah pertama di SMP Negeri 01 Sidomulyo pada tahun 2012. Sekolah menengah atas di SMA Negeri 01 Sidomulyo pada tahun 2015.

Pada tahun 2015 penulis terdaftar sebagai mahasiswi S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui Jalur SNMPTN. Selama menjadi mahasiswi, penulis ikut serta dalam organisasi Himpunan Mahasiswa Jurusan Matematika FMIPA Unila sebagai anggota. Pada tahun 2018, sebagai bentuk aplikasi bidang ilmu di dunia kerja, penulis telah melaksanakan Kerja Praktik (KP) selama 40 hari di kantor Badan Pusat Statistik Kabupaten Lampung Selatan. Pada tahun 2019, sebagai bentuk pengabdian kepada masyarakat, penulis telah melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Negeri Agung, Kecamatan Gunung Pelindung, Kabupaten Lampung Timur.

MOTTO

"Sesungguhnya Allah bersama orang-orang yang sabar"
(QS: Al-Baqarah:153)

" *The past can't be changed and the future is yet in your power*"
(Levi Ackerman)

"Akan banyak pelajaran yang akan diperoleh saat kita berani mengambil resiko"
(Levi Ackerman)

PERSEMBAHAN

Alhamdulillahirobbil'alamin,

Puji dan syukur tiada hentinya terpanjatkan kepada Allah SWT atas ridhonya sehingga penulis dapat menyelesaikan skripsi ini. Saya persembahkan karya ini untuk:

Kepada kedua orang tuaku, mami tersayang dan ayah kebanggaan yang selalu memberikan doa dan dukungan terus menerus kepadaku disaat mulai ingin menyerah. Kesabaran kalian dalam menghadapiku benar benar luar biasa.

Kepada dosen-dosen Pembimbing dan Pembahas yang telah sangat sabar dalam membimbing dan memberikan motivasi dan masukan ide ide yang membangun sehingga aku dapat menyelesaikan skripsi ini.

Sahabat tercinta, terimakasih atas kebersamaan, keceriaan, doa dan semangat yang telah diberikan.

Almamater kebanggaan, Universitas Lampung

SANWACANA

Puji dan syukur penulis haturkan kepada Allah SWT atas izin serta ridho-Nya dalam menyelesaikan skripsi yang berjudul "Estimasi Model Regresi Nonparametrik Menggunakan Estimator Nadaraya-Watson Dengan Fungsi Kernel *Triangle*". Shalawat serta salam tak lupa kepada Nabi Muhammad SAW yang telah menjadi suri tauladan yang baik bagi kita semua.

Penulis menyadari bahwa dalam penulisan skripsi ini tidak terlepas dari bimbingan, bantuan, dan kerjasama dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terimakasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku Dosen Pembimbing I, yang senantiasa selalu membimbing dan memberikan arahan, ide, kritik dan saran serta semangat kepada penulis selama proses pembuatan skripsi ini.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Dosen Pembimbing II, yang telah memberikan bimbingan serta saran yang membantu penulis dalam menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku Dosen Pembahas, terima kasih atas kesediaannya untuk membahas, memberikan saran dan kritik yang membangun dalam penyelesaian skripsi ini.

4. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku Dosen Pembimbing Akademik yang selalu memberikan bimbingan dalam menjalani perkuliahan.
5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Drs. Suratman, M.Sc., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh Dosen, Staf, dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Ibu, Bapak dan keluarga tercinta yang tak pernah berhenti memberi semangat, doa, nasihat dan kasih sayang serta pengorbanan yang tak tergantikan hingga penulis selalu kuat menjalani setiap rintangan yang ada di depan.
9. Sahabat-sahabat seperjuangan: Anisa, Mute, Salma, Fransiska, Riswanti, Meilinda, Purwanti, Neli, Della, Aulia Rahman, Ade, Bagus, Irmaningsih, Dita, Wahyu yang selalu menemani hari-hari penulis selama menjalani masa perkuliahan serta Dwi, Tri dan Fera teman perjuangan selama KKN.
10. Teman-temanku Matematika 2015, terimakasih telah memberikan warna dan keceriaan kepada penulis selama menjadi mahasiswi.
11. Almamater tercinta Universitas Lampung.
12. Seluruh pihak yang telah membantu yang tidak dapat disebutkan satu persatu.

Bandar Lampung, 06 Agustus 2019

Penulis,

Aulia Putri

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan metode analisis data yang menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor. Misalkan x adalah variabel prediktor dan y adalah variabel respon untuk n pengamatan berpasangan $\{(x_i, y_i)\}_{i=1}^{n}$, maka hubungan linear antara variabel prediktor dan variabel respon tersebut dapat dinyatakan sebagai berikut:

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

dengan ε_i adalah sisaan yang diasumsikan independen dengan mean nol dan variansi σ^2, serta $m(x_i)$ adalah fungsi regresi atau kurva regresi.

Pendekatan yang digunakan untuk mengestimasi fungsi regresi ada tiga, yaitu pendekatan parametrik, semiparametrik dan nonparametrik. Dalam pendekatan parametrik, bentuk hubungan antara variabel respon dan variabel prediktor diketahui atau diperkirakan dari bentuk kurva regresi, misalnya diasumsikan membentuk pola linear, kuadratik, eksponensial, dan polinomial. Apabila tidak terdapat informasi apapun terhadap bentuk fungsi serta tidak memenuhi asumsi normalitas dan homogenitas ragam dari galat data, maka digunakan pendekatan nonparametrik.

Pendekatan nonparametrik tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar. Pendugaan fungsi regresi nonparametrik dilakukan berdasarkan data menggunakan teknik pemulusan (*smoothing)* seperti histogram, penduga kernel, k-*Nearest Neighbor,* deret orthogonal, penduga spline, deret fourier dan wavelet. Teknik tersebut mempunyai keunggulan didalam mengestimasi parameter. Salah satu metode estimasi parameter adalah dengan menggunakan pendekatan kernel yang memiliki bentuk fleksibel dan perhitungan matematisnya mudah disesuaikan serta memiliki rata-rata kekonvergenan yang relatif cepat.

Ada beberapa jenis fungsi kernel, antara lain kernel *Uniform*, *Triangle*, Epanechnikov, *Gaussian*, *Kuadratik*, dan Cosinus. Dalam regresi kernal pemilihan parameter pemulus (*bandwidth*) jauh lebih penting dibandingkan dengan memilih fungsi kernel. Sehingga yang menjadi masalah dalam regresi kernel adalah pemilihan *bandwidth*, bukan pada pemilihan fungsi kernel. Terdapat beberapa kriteria pemilihan *bandwidth* yang telah diperkenalkan, seperti *Cross-Validation* (CV), *Generalized Cross Validation* (GCV), *Bayesian Information Criterion* (BIC), *Minimum Description Length* (MDL), *Akaike Information Criterion* (AIC) serta *Improved Akaike Information Criterion* (AICC).

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengestimasi kurva regresi nonparametrik menggunakan estimator Nadaraya-Watson dengan fungsi kernel *Triangle* pada data pertumbuhan balita di Desa Sidorejo Kec. Sidomulyo Lampung Selatan dan data simulasi.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mampu mengetahui dan memahami regresi nonparametrik menggunakan metode Nadaraya-Watson dengan fungsi kernel *Triangle* dengan metode pemilihan *bandwidth* optimal *Generalized Cross Validation* (GCV).
2. Penelitian ini dapat dijadikan sebagai bahan rujukan dan pengembangan pembelajaran statistika khususnya model regresi nonparametrik.
3. Mengembangkan wawasan keilmuan dan pengetahuan tentang regresi nonparametrik kernel metode Nadaraya-Watson.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi adalah suatu metode analisis data yang menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor. Misalkan x adalah variabel prediktor dan y adalah variabel respon untuk n data pengamatan berpasangan $\{(x_i, y_i)\}_{i=1}^{n}$, maka hubungan antara variabel prediktor dan variabel respon tersebut dapat dinyatakan sebagai berikut:

$$y_i = m(x_i) + \varepsilon_i \,, \; i = 1,2, \dots, n \qquad (2.1)$$

dengan ε_i adalah galat yang diasumsikan independen dengan mean 0 dan variansi σ^2 (konstan), $m(x_i)$ disebut sebagai fungsi regresi atau kurva regresi (Hardle, 1994).

Ada dua pendekatan yang dapat digunakan untuk mengestimasi fungsi regresi atau kurva regresi, yaitu secara parametrik dan nonparametrik. Dalam pendekatan parametrik, bentuk hubungan antara variabel respon dan variabel prediktor diketahui atau diperkirakan dari bentuk kurva regresi. Sedangkan pada regresi nonparametrik, bentuk kurva tidak dapat langsung diketahui atau diperkirakan (Netter, *et al*., 1997)

2.2 Regresi Nonparametrik

Pendekatan regresi nonparametrik digunakan bila tidak terdapat informasi mengenai bentuk fungsi regresi nonparametrik $m(x_i)$ dan tidak bergantung pada asumsi bentuk kurva tertentu. Data diharapkan mencari sendiri bentuk estimasinya sehingga memiliki fleksibilitas yang tinggi. Kurva regresi hanya diasumsikan termuat dalam suatu ruang fungsi yang berdimensi tak hingga dan merupakan fungsi mulus (*smooth*). Pendugaan fungsi regresi nonparametrik $m(x_i)$ dilakukan berdasarkan data menggunakan teknik pemulusan tertentu, yaitu penduga histogram, kernel, deret orthogonal, spline, k-NN, deret fourier, dan wavelet (Eubank, 1998).

Meskipun begitu, regresi nonparametrik bukannya tanpa kekurangan. Kekurangan atau kelemahan dari regresi nonparametrik menurut Hollander, M., *et al.* (2014) antara lain:

1. Metode ini dianggap mengorbankan terlalu banyak informasi dari data yang didapat.
2. Metode ini dianggap tidak efisien daripada metode parametrik pada data tertentu.
3. Hanya dapat digunakan pada data dengan n yang kecil.
4. Metode nonparametrik relatif tidak sensitif pada pencilan data.

2.3 Uji Normalitas

Menurut Montgomery & Peck (1992) asumsi kenormalan seringkali tidak terpenuhi karena adanya pengamatan pencilan yang memberikan pengaruh besar terhadap estimasi parameter model. Pengujian asumsi residual normal $(0, \sigma^2)$ dapat dilakukan melalui uji *Kolmogorov-Smirnov*. Hipotesis yang digunakan adalah:

$H_0 : F_0(x) = F(x)$ (Residual berdistribusi Normal $(0, \sigma^2)$)

$H_1 : F_0(x) \neq F(x)$ (Residual tidak berdistribusi Normal $(0, \sigma^2)$)

dengan statistik uji:

$$D = maks|F_0(x) - S_N(x)| \tag{2.2}$$

dimana $F_0(x)$ adalah fungsi distribusi kumulatif teoritis sedangkan $S_N(x) = i/n$ merupakan fungsi peluang kumulatif pengamatan dari suatu sampel random dengan i adalah pengamatan dan n adalah jumlah pengamatan. Pengambilan keputusan adalah H_0 ditolak jika $|D| > q_{(1-\alpha)}$ dimana q adalah nilai berdasarkan tabel *Kolmogorov-Smirnov*. Selain itu juga dapat melalui $p - value$, dimana H_0 ditolak jika $p - value$ kurang dari α.

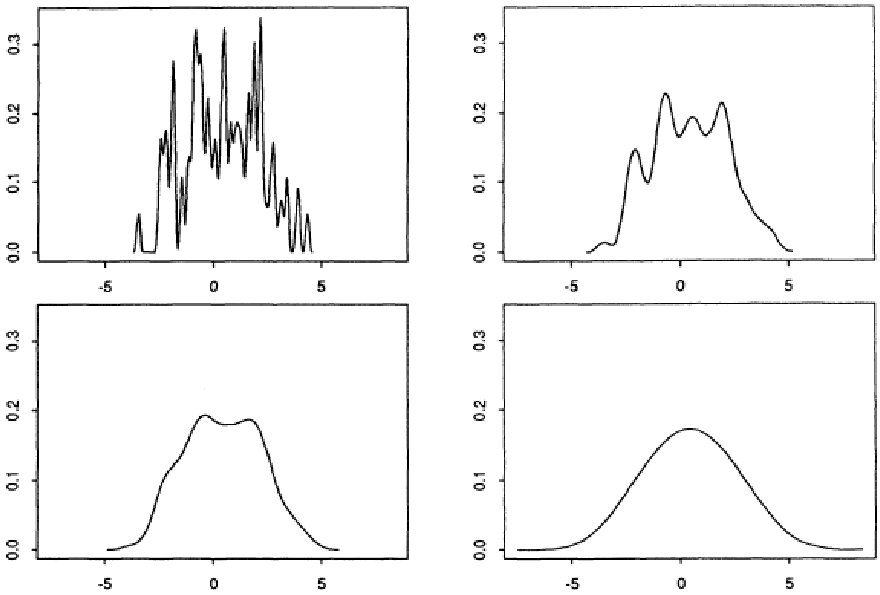
2.4 Uji Distribusi

Data yang diperoleh dari penelitian harus dilakukan uji distribusi. Pengujian distribusi data dilakukan untuk mengetahui jenis distribusi data yang diperoleh. Salah satu uji yang dapat digunakan untuk menguji kecocokan antara distribusi frekuensi data dengan hasil model model yang dikembangkan adalah uji *Kolmogorov-Smirnov* (Alfigari, 1997).

Uji *Kolmogorov-Smirnov* digunakan untuk menguji apakah distribusi data sampel yang teramati sesuai dengan distribusi teoritis tertentu atau tidak. Uji *Kolmogorov-Smirnov* beranggapan bahwa distribusi data yang diuji bersifat kontinu dan sampel diambil dari populasi secara acak. Prinsip uji *Kolmogorov-Smirnov* adalah menghitung selisih antara fungsi distribusi frekuensi komulatif sampel $(F_s(X))$ dan fungsi distribusi komulatif teoretis $(F_t(X))$ pada masing masing interval kelas.

2.5 Penduga Kernel

Penduga kernel diperkenalkan oleh Rosenblatt pada tahun 1956 dalam Hardle (1991) yang merupakan pengembangan dari estimator histogram. Rosenblatt mengusulkan menempatkan *smoothing* kernel di setiap pengamatan. Kernel memiliki parameter pemulus yang mengatur tingkat kehalusan kurva dugaan kernel yang disebut *bandwidth* (h). Pemilihan h akan memengaruhi hasil *smoothing* kernel. Nilai h yang semakin kecil akan menyebabkan bentuk kurva semakin kasar dan sebaliknya semakin besar nilai h akan menyebabkan kurva semakin mulus. Hal ini seperti yang tersaji seperti Gambar 1.



Gambar 1. Estimasi kernel dengan nilai $h = 0.2, 0.8, 1.4$ dan 4.

Hardle (1991) menyatakan bahwa penduga kernel memiliki beberapa keuntungan diantaranya:

1. Penduga kernel mempunyai bentuk yang fleksibel serta mudah dalam perhitungan matematis.
2. Penduga kernel mempunyai bentuk kekonvergenan yang relatif cepat.

Secara umum kernel K dengan parameter pemulus h didefinisikan sebagai:

$$K_h(x) = \frac{1}{h}K(\frac{x}{h}) \qquad \text{untuk } -\infty < x < \infty \text{ dan } h > 0 \tag{2.3}$$

Serta memenuhi:

(1) $K(x) \geq 0$, untuk semua x

(2) $\int_{-\infty}^{\infty} K(x)dx = 1$

(3) $\int_{-\infty}^{\infty} t^2 K(x)dx = \sigma^2 > 0$

(4) $\int_{-\infty}^{\infty} tK\,(x)dx = 0$

Sementara penduga densitas kernel untuk fungsi densitas f(x) didefinisikan sebagai:

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

$$= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \qquad (2.4)$$

Beberapa jenis fungsi kernel antara lain:

1. Kernel Uniform : $K(x) = \frac{1}{2}$; $|x| \leq 1{,}0$ selainnya
2. Kernel Triangel : $K(x) = (1 - |x|)$; $|x| \leq 1{,}0$ selainnya
3. Kernel Epanechikov : $K(x) = \frac{3}{4}(1 - x^2)$; $|x| \leq 1{,}0$ selainnya
4. Kernel Kuartik : $K(x) = \frac{15}{16}(1 - x^2)^2$; $|x| \leq 1{,}0$ selainnya
5. Kernel Triweight : $K(x) = \frac{35}{32}(1 - x^2)^3$; $|x| \leq 1{,}0$ selainnya
6. Kernel Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$; $-\infty < x < \infty$
7. Kernel Cosinus : $K(x) = \frac{\pi}{4}\cos\left(\frac{\pi}{2}x\right)$; $|x| \leq 1{,}0$ selainnya

Dari persamaan (2.4) terlihat bahwa $\hat{f}_h(x)$ penduga densitas kernel tergantung pada dua parameter yaitu fungsi kernel (K) dan parameter h. Bentuk bobot kernel ditentukan oleh fungsi kernel (K), sedangkan ukuran bobotnya ditentukan oleh parameter pemulus h yang disebut *bandwidth* (Wand & Jones, 1995).

Dalam regresi kernel pemilihan h jauh lebih penting dibandingkan dengan pemilihan fungsi kernel. Hal ini disebabkan penggunaan fungsi kernel yang berbeda dengan nilai h optimal menghasilkan estimasi kurva regresi yang hampir sama. Permasalahan dalam regresi kernel adalah pemilihan h, bukan pada pemilihan fungsi kernel. Fungsi kernel yang umum digunakan adalah kernel

Gaussian dan kernel Epanechnicov. Kernel Triangle sering digunakan karena lebih mudah dan cepat dalam perhitungan (Sukarsa & Srinadi, 2012).

2.6 Penduga Nadaraya-Watson

Menurut Hardle (1991), jika terdapat n data pengamatan $\{(X_i, Y_i)\}_{i=1}^{n}$ yang memenuhi persamaan dimana $X_i \in R$ dan $Y_i \in R$, maka penduga $m(x_i)$ adalah:

$$\widehat{m}(x) = E(Y|X = x) = \int \frac{yf(X,Y)}{f(X=x)} dy \tag{2.5}$$

Penyebut pada persamaan (2.5) diduga dengan menggunakan penduga densitas kernel sebagai berikut:

$$f_h(x) = \frac{1}{n} \sum_{i=1}^{n} (x - X_i)$$

Fungsi densitas peluang bersama diduga dengan perkalian kernel, yaitu:

$$\acute{f}_{h_1,h_2}(x, y) = \frac{1}{n} \sum_{i=1}^{n} K_{h_1}(x - X_i) K_{h_2}(y - Y_i)$$

Sehingga, pembilang dari penduga Nadaraya-Watson menjadi:

$$\begin{aligned}
\int y \acute{f}_{h_1,h_2}(x, y) dy &= n^{-1} \sum_{i=1}^{n} K_{h_1}(x - X_i) \int y K_{h_2}(y - Y_i) dy \\
&= n^{-1} \sum_{i=1}^{n} K_{h_1}(x - X_i) \int \frac{y}{h_2} K\left(\frac{y - Y_i}{h_2}\right) dy \\
&= n^{-1} \sum_{i=1}^{n} K_{h_1}(x - X_i) \int (sh_2 + Y_i) K(s)\, ds \\
&= n^{-1} \sum_{i=1}^{n} K_{h_1}(x - X_i) Y_i
\end{aligned}$$

Dengan demikian penduga Nadaraya-Watson dapat ditulis:

$$\widehat{m}(x_i)=\frac{\frac{1}{n}\sum_{j=1}^{n}K_h(x_i-X_j)Y_i}{\frac{1}{n}\sum_{k=1}^{n}K_h(x_i-X_k)}$$

$$\widehat{m}(x_i)=\frac{\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{x_i-X_j}{h}\right)Y_i}{\frac{1}{nh}\sum_{k=1}^{n}K\left(\frac{x_i-X_k}{h}\right)}$$

$$\widehat{m}(x_i)=\frac{\sum_{j=1}^{n}K\left(\frac{x_i-X_j}{h}\right)Y_i}{\sum_{k=1}^{n}K\left(\frac{x_i-X_k}{h}\right)} \qquad (2.6)$$

$$\widehat{m}(x_i)=\sum_{i=1}^{n}W_{ij}(x_i)Y_i$$

Sehingga, $\hat{Y}=WY$, dimana

$$W_{ij}(x_i)=\frac{K\left(\frac{x_i-X_j}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{x_i-X_j}{h}\right)}$$

Matriks W_{ij} disebut juga dengan *Hat Matrix* dari penduga $m(x_i)$. Persamaan (2.6) ditemukan oleh Nadaraya & Watson (1964) dalam Hardle (1991), sehingga disebut estimator Nadaraya-Watson.

Dalam penelitian ini, akan digunakan fungsi kernel *Triangle* yang didefinisikan sebagai berikut:

$$K(x)=(1-|x|)\ ; \quad |x|\le 1\ ,0 \text{ selainnya.} \qquad (2.7)$$

Dengan mensubtitusikan persamaan (2.9) kedalam persamaan $\widehat{m}(x_i)$ maka penduga Nadaraya-Watson menjadi:

$$\widehat{m}(x_i)=\frac{\sum_{j=1}^{n}\left(\left(1-\left|\frac{x_i-X_j}{h}\right|\right)Yi\right)}{\sum_{k=1}^{n}\left(1-\left|\frac{x_i-X_k}{h}\right|\right)} \qquad (2.8)$$

Pengaruh fungsi kernel kurang signifikan dibandingkan dengan pengaruh h. Nilai-nilai ekstrik dari h mengakibatkan:

- Jika $h \to 0$, maka untuk $x = x_{i,}$ $m(x_i) \to \frac{K(0)}{K(0)} Y_i = Y_i$

Jadi, jika h sangat kecil, estimator akan menuju ke data.

- Jika $h \to \infty$ maka $K\left(\frac{x-X_i}{h}\right) \to K(0)$, akibatnya

$$m(x_i) \to \frac{n^{-1}\sum_{i=1}^{n} K(0)Y_i}{n^{-1}\sum_{i=1}^{n} K(0)} = \frac{n^{-1}K(0)\sum_{i=1}^{n} Y_i}{n^{-1}(nK(0))} = n^{-1}\sum_{i=1}^{n} Y_i$$

Jadi, jika h sangat besar, estimator akan sangat mulus dan menuju rata-rata dari variabel respon. Semakin kecil nilai h, maka grafik akan semakin kurang mulus namun memiliki bias yang kecil. Sebaliknya semakin besar nilai h, maka grafik akan sangat mulus tetapi memiliki bias yang besar. Karena tujuan estimasi kernel adalah memperoleh kurva yang mulus namun memiliki nilai MSE yang tidak terlalu besar, perlu dipilih nilai h optimal untuk mendapatkan grafik optimal. Salah satu cara memilih parameter pemulus optimal adalah dengan menggunakan metode *Generalized Cross Validation* (GCV).

2.7 Pemilihan *Bandwidth* Optimum

Menurut Hardle (1991), *bandwidth* (h) adalah parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan sangat fluktuatif, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data. Tujuan estimasi kurva tidak hanya untuk memperoleh kurva yang mulus tetapi juga memiliki tingkat kesalahan yang tidak terlalu besar. Berdasarkan hal itu perlu dipilih nilai *bandwidth* optimal sehingga didapatkan kurva yang mulus dengan kesalahan yang minimum.

Salah satu metode untuk mendapatkan h optimal diperoleh dengan menggunakan kriteria *Generalized Cross Validation* (GCV), yang didefinisikan sebagai berikut :

$$GCV\ (h) = \frac{MSE(h)}{\left[1 - \frac{trace(W)}{n}\right]^2} \tag{2.9}$$

dengan

$$MSE\ (h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - m(x_i))^2 \tag{2.10}$$

dan W adalah hat matriks berukuran $n \times n$ yang memenuhi

$$[m_h(x_1), m_h(x_2), \dots, m_h(x_n)|^t = WY \tag{2.11}$$

serta $trace(W)$ adalah jumlah elemen pada diagonal utama, yaitu diagonal dari kiri atas ke kanan bawah dari matriks W_{ij}. Nilai h optimal akan diperoleh jika nilai *Generalized Cross Validation* minimal (Craven & Wahba, 1979).

Jika data tidak simetrik dan unimodal, maka lebar jendela yang mengoptimalkan IMSE (*Integrated Mean Square Error*) diberikan oleh Silverman (1986) seperti rumus berikut:

$$h_{opt} = 0{,}9\left\{S; \frac{IQR}{1{,}34}\right\} n^{-1/5}$$

dengan S adalah simpangan baku, IQR disebut juga *Inter Quartil Range* (Jangkauan antar quartil) dan n merupakan banyaknya data pengamatan.

2.8 Ukuran Kebaikan *Bandwidth*

Kebaikan suatu penduga dapat dilihat dari tingkat kesalahannya, semakin kecil tingkat kesalahan semakin baik estimasinya. Menurut Chatterjee & Hadi (2006), salah satu kriteria untuk menentukan estimator terbaik dalam model regresi adalah nilai *R-Square* (R^2). *R-square* didefinisikan sebagai:

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT} \tag{2.12}$$

dengan $0 \leq R^2 \leq 1$, dimana R^2 makin dekat dengan 1 makin baik estimasinya, dan sebaliknya, makin dekat dengan 0 makin jelek estimasinya. JKT (Jumlah Kuadrat Total) merupakan jumlah kuadrat simpangan dari rata-rata variabel respon, JKR (Jumlah Kuadrat Regresi) merupakan jumlah kuadrat simpangan hasil dugaan dengan rata-rata variabel respon y, dan JKG (Jumlah Kuadrat Galat) mengukur residual dalam prediksi. Jadi dapat ditulis sebagai:

$$JKT = JKR + JKG$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.13}$$

dengan y_i adalah data variabel respon ke-i, $\bar{y}$ adalah rata-rata variabel respon, sedangkan $\hat{y}_i$ adalah nilai hasil dugaan variabel respon ke-i.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2018/2019 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data pertumbuhan balita di Posyandu Desa Sidorejo Kec. Sidomulyo, Lampung Selatan pada bulan Maret 2018 dengan variabel respon (y) yaitu berat badan balita dan variabel prediktor (x) yaitu usia yang diimplementasikan pada data simulasi menggunakan program aplikasi R3.5.2 dengan kondisi sebaran galat $\varepsilon \sim Uniform(-9{,}40;\ 3{,}10)$, sebaran variabel prediktor $x \sim Weibull\ (2{,}07014;\ 12{,}04575)$ pada $y = x + \varepsilon$.

3.3 Metode Penelitian

Pada penelitian ini model regresi nonparametrik diestimasi menggunakan metode Nadaraya-Watson dengan fungsi kernel *Triangle*. Program yang digunakan dalam penelitian ini adalah program R3.5.2 . Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Membuat diagram pencar antara variabel respon dan variabel prediktor pada data pertumbuhan balita.
2. Melakukan uji normalitas data.
3. Menentukan *bandwidth* optimal menggunakan fungsi kernel *Triangle* dan ditentukan dengan GCV minimum pada data pertumbuhan balita.
4. Mengukur kebaikan *bandwidth* optimal berdasarkan nilai *R-square*.
5. Menentukan kurva dugaan regresi berdasarkan nilai *bandwidth* optimal dengan metode Nadaraya-Watson pada data pertumbuhan balita.
6. Membangkitkan data sebanyak $n = 30$ dengan sebaran $x \sim Weibull\ (2{,}07014;\ 12{,}04575)$, $\varepsilon \sim Uniform(-9{,}40;\ 3{,}10)$ dan $y = x + \varepsilon$.
7. Membuat diagram pencar antara variabel respon dan variabel prediktor pada data simulasi.
8. Menentukan *bandwidth* optimal menggunakan fungsi kernel *Triangle* dan ditentukan dengan GCV minimum pada data simulasi.
9. Mengukur kebaikan *bandwidth* optimal berdasarkan nilai *R-square*.
10. Menentukan kurva dugaan regresi berdasarkan nilai *bandwidth* optimal dengan metode Nadaraya-Watson pada data simulasi.
11. Membuat kesimpulan.

V. KESIMPULAN

Berdasarkan hasil penelitian diperoleh kesimpulan bahwa estimasi model regresi nonparametrik dengan fungsi kernel *Triangle* pada data pertumbuhan balita di Posyandu Desa Sidorejo diperoleh estimasi kurva regresi yang optimal pada *bandwidth* 5,0 dengan nilai GCV $1{,}363146$. Sementara estimasi model regresi nonparametrik dengan fungsi kernel *Triangle* pada data simulasi (bangkitan) dengan kondisi $\varepsilon \sim Uniform(-9{,}40;\ 3{,}10)$, $x \sim Weibull\ (2{,}07014;\ 12{,}04575)$ dan $y = x + \ \varepsilon$ diperoleh estimasi kurva regresi yang optimal pada *bandwidth* 5,0 dengan nilai GCV $1{,}867415$.

DAFTAR PUSTAKA

Algifari. 1997. *Analisis Statistik untuk Bisnis dengan Regresi, Korelasi dan Nonparametrik.* STIE-YKPN, Yogyakarta.

Chatterjee, S. & Hadi, A.S. 2006. *Regression Analysis by Example.* Jhon Wiley & Sons, Inc., New Jersey.

Craven, P. & Wahba, G. 1979. Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numer Math University of Wisconsin.* **31**: 377-403.

Eubank, R. 1998. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.

Fitriani, dkk. 2015. Estimasi Model Regresi Semiparametrik Menggunakan Estimator Kernel Uniform. *Jurnal Matematika*. **4**(4): 176-180.

Hardle, W. 1991. *Smoothing Techniques with Implementation in S*. Cambridge University Press, New York.

Hardle, W. 1994. *Applied Nonparametric Regression*. Cambridge University Press, New York.

Hollander, M., *et al*. 2014. *Nonparametric Statistical Methods*. John Wiley & Sons, Inc., New Jersey

Montgomery, D.C., & Peck, E.A. 1992. *Introduction to Linear Regression Analysis.* John Wiley & Sons, Inc., Toronto.

Netter, J., *et al*. 1997. *Model Linear Terapan Analisis Regresi Linier Sederhana*. Diterjemahkan oleh Bambang Sumantri. Jurusan Statistika FMIPA IPB, Bogor.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Sukarsa, I.K.G. & Srinadi, I.G.A.M. 2012. Estimator Kernel dalam Model Regresi Nonparametrik. *Jurnal Matematika*. **2**(1): 19-30.

Wand, M.P. & Jones, M.C. 1995. *Kernel Smoothing*. Chapman & Hall, New York.