

SKRIPSI

**ESTIMASI MODEL REGRESI NONPARAMETRIK
MENGUNAKAN ESTIMATOR *NADARAYA-WATSON*
DENGAN FUNGSI KERNEL *EPANECHNIKOV***

Disusun dan diajukan oleh

NUR LIA

H051171512



**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR**

2021

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

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Statistika Departemen Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

NUR LIA

H051171512

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama : Nur Lia
NIM : H051171512
Program Studi : Statistika
Jenjang : Sarjana (S1)

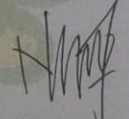
Menyatakan dengan ini bahwa karya tulisan saya berjudul

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

Adalah karya tulisan saya sendiri dan bukan merupakan pengambil alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 2 Juni 2021



NUR LIA

NIM. H051171512

LEMBAR PENGESAHAN (TUGAS AKHIR)

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

Disusun dan diajukan oleh

NUR LIA

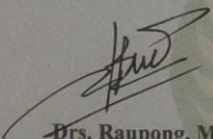
H051171512

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 2 Juni 2021
dan dinyatakan telah memenuhi syarat kelulusan.

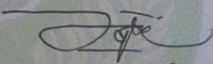
Menyetujui,

Pembimbing Utama,

Pembimbing Pertama,

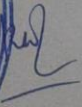

Drs. Raupong, M.Si.

NIP. 19621015 198810 1 001


Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19650519 1999303 2002




Dr. Nur H. Samusi, S.Si., M.Si.


NIP. 19720117 199703 2002

ESTIMASI MODEL REGRESI NONPARAMETRIK
MENGUNAKAN ESTIMATOR *NADARAYA-WATSON*
DENGAN FUNGSI KERNEL *EPANECHNIKOV*

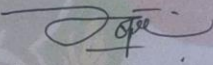
Disetujui Oleh:

Pembimbing Utama,

Pembimbing Pertama,



Drs. Raupong, M.Si.

NIP. 19621015 198810 1 001


Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19650519 1999303 2002




Dr. Nurul Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

Pada Tanggal : 2 Juni 2021

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, karunia, serta hidayah-Nya kepada penulis sehingga skripsi ini dapat diselesaikan dengan baik. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW dan kepada para keluarga serta sahabat beliau. Skripsi ini berjudul “**Estimasi Model regresi Nonparametrik Menggunakan Estimator Nadaraya-Watson dengan Fungsi Kernel Epanechnikov**” yang disusun untuk memenuhi salah satu syarat akademik guna meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa penulisan skripsi ini tidak mungkin sukses tanpa adanya bantuan dari berbagai pihak, baik bantuan moril maupun materiil. Oleh karena itu, penulis dengan rasa rendah hati mengucapkan terima kasih yang sebesar-besarnya kepada keluarga terutama Papa tercinta **Bahtiar**, Mama **Jumasang**, dan Kakak **Nur Ilma, S.TP** atas semua kasih sayang, kesabaran, perhatian, dan dukungan yang diberikan. Demikian pula penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu, MA**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. Bapak **Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku Ketua Departemen Statistika FMIPA UNHAS dan Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku Sekretaris Departemen Statistika dan **segenap dosen pengajar**, terima kasih telah memberikan ilmu kepada penulis serta **staf Departemen Statistika** terima kasih telah membantu dalam pengurusan persuratan.
4. Bapak **Drs. Raupong, M.Si.** selaku Dosen Pembimbing Utama sekaligus Pembimbing Akademik dan Ibu **Dr. Anna Islamiyati, S.Si., M.Si.** selaku Dosen Pembimbing Pertama yang telah meluangkan waktu ditengah berbagai

kesibukan untuk membimbing dan membagi ilmu serta memberikan dukungan kepada penulis dalam menyelesaikan tugas akhir ini.

5. Ibu **Sitti Sahriman, S.Si., M.Si.** dan Bapak **Dr. La Podje Talangko, M.Si.** selaku penguji, terima kasih atas segala masukan serta nasehat yang telah diberikan kepada penulis selama penyusunan tugas akhir ini.
6. Teman-teman **Statistika 2017** terima kasih untuk segala kebersamaan, kerja sama, kenangan serta dukungan yang telah diberikan kepada penulis.
7. **Hana Meisaria Retno Permataliyanti** terima kasih karena selalu menemani dalam penyusunan, pengurusan persuratan dan berkas serta menjadi tempat bertukar pikiran selama penyelesaian tugas akhir ini. Untuk **Dwi Auliyah** dan **Fadhillah Oktavia Nabir** terima kasih untuk segala kenangan dan telah membuat hari-hari perkuliahan menjadi lebih berwarna. Untuk **iklil dan shafwan** terima kasih telah menjadi tempat bertukar pikiran selama penyelesaian tugas akhir ini.
8. Teman-teman **KKN Gelombang 104 Kabupaten Kepulauan Selayar** terima kasih untuk segala dukungan dan kenangan kebersamaannya.
9. Untuk **Desty dan Ifah** terima kasih selalu ada setiap kali penulis butuh bantuan dan semangat yang diberikan kepada penulis.
10. Semua pihak yang telah memberikan bantuan dan dukungan secara langsung maupun tidak langsung yang tidak sempat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah SWT.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Makassar, 2 Juni 2021

Nur Lia

ABSTRAK

Analisis regresi merupakan suatu metode yang digunakan untuk memodelkan hubungan antara variabel respon (y) dengan variabel prediktor (x). Untuk data rata-rata harian posisi *sunspot* menggunakan pendekatan regresi nonparametrik. Terdapat beberapa metode untuk mengestimasi model pada regresi nonparametrik salah satunya adalah estimator kernel. Estimator kernel merupakan salah satu metode pendekatan terhadap fungsi densitas yang belum diketahui dengan menggunakan fungsi kernel. Hal yang paling penting dalam pendekatan kernel adalah pemilihan *bandwidth*. *Bandwidth* berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Salah satu kriteria yang dapat digunakan untuk memilih *bandwidth* yang optimum adalah *Generalized Cross Validation (GCV)*. Nilai GCV minimum menunjukkan nilai *bandwidth* yang optimum. Penelitian ini bertujuan untuk memperoleh parameter model regresi nonparametrik menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov* untuk data rata-rata harian bilangan *sunspot* pada tahun 2007. Berdasarkan hasil analisis dan pembahasan dapat diketahui bahwa estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov* baik diterapkan pada data rata-rata harian bilangan *sunspot* pada tahun 2007. Estimator ini menghasilkan kurva regresi yang relatif sama dengan kurva data asli. Diperoleh nilai *bandwidth* optimum adalah 1.0 dengan nilai GCV minimum sebesar $2.03E-31$ serta nilai MSE yang relatif kecil sebesar $2.02E-31$

Kata kunci : Regresi Nonparametrik, Estimator Kernel *Nadaraya-Watson*, Fungsi Kernel *Epanechnikov*, *Bandwidth*, *Generalized Cross Validation*, Posisi *Sunspot*.

ABSTRACT

Regression analysis is a method used to model the relationship between the response variable (y) and the predictor variable (x). For the daily average sunspot position data using a nonparametric regression approach. There are several methods for estimating the model in nonparametric regression, one of which is the kernel estimator. The kernel estimator is a method of approaching the unknown density function by using a kernel function. The most important thing in the kernel approach is bandwidth selection. Bandwidth serves to control the smoothness of the estimated curve. One of the criteria that can be used to select the optimum bandwidth is Generalized Cross Validation (GCV). The minimum GCV value indicates the optimum bandwidth value. This study aims to obtain nonparametric regression model parameters using the Nadaraya-Watson estimator with the Epanechnikov kernel function for the daily average sunspot number in 2007. Based on the results of the analysis and discussion, it can be seen that the Nadaraya-Watson estimator with the Epanechnikov kernel function is well applied to the data daily average of sunspot numbers in 2007. This estimator produces a regression curve that is relatively the same as the original data curve. The optimum bandwidth value is 1.0 with a minimum GCV value of $2.03E-31$ and a relatively small MSE value of $2.02E-31$

Keywords : Nonparametric Regression, *Nadaraya-Watson* Kernel Estimator, *Epanechnikov* Kernel Function, *Bandwidth*, *Generalized Cross Validation*, *Sunspot* Position.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERNYATAAN KEASLIAN	iii
LEMBAR PENGESAHAN (TUGAS AKHIR)	iv
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
BAB II. LANDASAN TEORI	5
2.1 Analisis Regresi	5
2.2 Regresi Nonparametrik	5
2.3 Estimator Kernel	6
2.4 Fungsi Kernel	7
2.5 Estimator <i>Nadaraya-Watson</i>	8
2.6 Pemilihan <i>Bandwidth</i> Optimum	9
2.7 Matahari	11
2.8 <i>Sunspot</i> Matahari	12
BAB III. METODOLOGI PENELITIAN	14
3.1 Sumber Data	14
3.2 Variabel Penelitian	14
3.3 Metode Analisis	14

3.3.1 Langkah-langkah Mendapatkan Kurva Estimasi	15
BAB IV. HASIL DAN PEMBAHASAN	17
4.1 Deskripsi Data	17
4.2 Analisis Data	19
BAB V. KESIMPULAN DAN SARAN	29
5.1 Kesimpulan	29
5.2 Saran	29
DAFTAR PUSTAKA	31
LAMPIRAN	32

DAFTAR GAMBAR

Gambar 4.1 Posisi <i>Sunspot</i> di Lintang Selatan	17
Gambar 4.2 Kurva Regresi Kernel <i>Epanechnikov</i> Terhadap Posisi <i>Sunspot</i> di Lintang Selatan dengan $h = 0.1$	20
Gambar 4.3 Kurva Error Estimasi Regresi Kernel <i>Epanechnikov</i> dengan $h = 0.1$	21
Gambar 4.4 Kurva Regresi Kernel <i>Epanechnikov</i> Terhadap Posisi <i>Sunspot</i> di Lintang Selatan dengan $h = 2.0$	22
Gambar 4.5 Kurva Error Estimasi Regresi Kernel <i>Epanechnikov</i> dengan $h = 2.0$	22
Gambar 4.6 Kurva Regresi Kernel <i>Epanechnikov</i> Terhadap Posisi <i>Sunspot</i> di Lintang Selatan dengan $h = 1.0$	24
Gambar 4.7 Kurva Error Estimasi Regresi Kernel <i>Epanechnikov</i> dengan $h = 1.0$	24
Gambar 4.8 Kurva Regresi Kernel <i>Epanechnikov</i> Terhadap Posisi <i>Sunspot</i> di Lintang Selatan dengan $h = 0.9$	25
Gambar 4.9 Kurva Error Estimasi Regresi Kernel <i>Epanechnikov</i> dengan $h = 0.9$	26
Gambar 4.10 Kurva Regresi Kernel <i>Epanechnikov</i> Terhadap Posisi <i>Sunspot</i> di Lintang Selatan dengan $h = 1.1$	27
Gambar 4.11 Kurva Error Estimasi Regresi Kernel <i>Epanechnikov</i> dengan $h = 1.1$	27

DAFTAR TABEL

Tabel 4.1 Statistika Deskriptif Data Rata-Rata Harian Posisi <i>Sunspot</i> di Lintang Selatan Tahun 2007	18
Tabel 4.2 Nilai GCV dari Interval <i>Bandwidth</i> 0.1 sampai 2.0	23
Tabel 4.3 Nilai MSE dari <i>Bandwidth</i> 0.9, 1.0, dan 1.1	28

DAFTAR LAMPIRAN

Lampiran 1. Data Rata-rata Harian Posisi <i>Sunspot</i> di Lintang Selatan Pada Tahun 2007	33
Lampiran 2. Source Code dan Output Grafik Data Rata-Rata Harian Posisi <i>Sunspot</i> di Lintang Selatan	36
Lampiran 3. Nilai MSE, RMSE, dan MAD untuk <i>Bandwidth</i> 0.1 sampai 2.0 ..	47

BAB I

PENDAHULUAN

1.1. Latar Belakang

Matahari merupakan salah satu benda terpenting yang berada di pusat tata surya. Ukuran matahari sangatlah besar, diameternya mencapai 1.400 km dengan massa sekitar 2 kali 1.030 kilogram. Matahari memiliki bentuk yang tidak sepenuhnya bulat dan berbentuk bola gas pijar dengan zat plasma yang terdiri dari elektron dan proton. Suhu pada permukaan bola gas ini sangat panas mencapai 5.800 derajat Kelvin dan pusatnya memiliki suhu 15 juta derajat Kelvin. Matahari memancarkan panas dan cahaya untuk kelangsungan makhluk hidup di bumi. Tanpa adanya matahari, bumi ini akan menjadi tandus dan semua tempat akan menjadi gurun yang beku (Kaufman, 1978).

Sebagian besar cahaya matahari berasal dari fotosfer. Lapisan fotosfer merupakan salah satu lapisan matahari yang tampak dari bumi. Salah satu fenomena matahari yang mudah dikenal adalah munculnya bintik matahari (*sunspot*). *Sunspot* adalah suatu daerah aktif di lapisan fotosfer matahari yang mempunyai temperatur lebih rendah dari pada sekelilingnya sehingga tampak gelap. *Sunspot* terjadi karena adanya aktivitas medan magnet yang kuat pada permukaan matahari yang disebabkan oleh rotasi diferensial. Garis-garis medan magnet pada permukaan matahari akan keluar dan menembus lapisan fotosfer. Aktivitas sunspot mempunyai pola sebaran unik yang berulang. Perulangan inilah yang disebut siklus matahari. Matahari akan mengalami perubahan aktivitas yang ditandai dengan perubahan jumlah sunspot. Banyaknya jumlah sunspot menunjukkan tingkat keaktifan matahari. Makin banyak bintik yang muncul menandakan bahwa matahari makin aktif (Volvacea, 2012).

Pola sebaran bilangan sunspot diasumsikan tidak mengikuti pola sebaran data normal. Untuk data yang tidak diketahui pola sebarannya digunakan metode regresi nonparametrik untuk mendekati pola sebaran data aslinya. Estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan dengan menggunakan teknik smoothing (Eubank, 1998). Terdapat beberapa teknik

smoothing dalam model regresi nonparametrik antara lain *histogram*, *estimator kernel*, *deret orthogonal*, *estimator spline*, *k-NN*, *deret fourier*, dan *wavelet*. Salah satu diantara beberapa teknik dalam regresi nonparametrik yang dapat digunakan sangat cepat dan mudah perhitungannya adalah regresi kernel (Hardle, 1990). Regresi kernel memiliki keunggulan bentuknya yang lebih fleksibel. Pada regresi kernel juga terdapat estimator yang paling banyak digunakan untuk mengestimasi fungsi regresi yaitu estimator *Nadaraya-Watson*.

Masalah terpenting dalam penggunaan regresi kernel adalah pemilihan *bandwidth* yang optimal. *Bandwidth* dari kernel adalah parameter pemulus yang menunjukkan pengaruh yang kuat pada perkiraan yang dihasilkan. Selain itu *bandwidth* berfungsi untuk mengontrol kemulusan kurva yang diestimasi. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*), sebaliknya jika *bandwidth* terlalu besar maka akan menghasilkan estimasi kurva yang sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran data aslinya. Jadi *bandwidth* yang dipilih harus optimal. Pemilihan *bandwidth* yang optimal dilakukan dengan cara memilih nilai *bandwidth* yang minimum. Salah satu cara untuk mendapatkan *bandwidth* yang optimal adalah dengan *Generalized Cross Validation* (GCV). Selain *bandwidth* terdapat beberapa jenis fungsi kernel, antara lain kernel *uniform*, *triangle*, *Epanechnikov*, *Quartic*, *Gaussian*, dan *Cosinus* (Hardle, 1990).

Menurut Komang dan Gusti (2012) fungsi kernel yang umum digunakan adalah kernel *epanechnikov* dan kernel *triangle* karena mudah dan cepat dalam perhitungan. Hardle (1990) juga menyatakan bahwa kernel *Epanechnikov* merupakan kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding kernel yang lainnya, sedangkan kernel *triangle* dalam penelitian yang dilakukan oleh Puspitasari, dkk (2012) merupakan fungsi kernel yang memiliki nilai MSE yang lebih kecil dari fungsi kernel yang lain.

Penelitian terkait aplikasi regresi nonparametrik kernel telah dilakukan oleh beberapa peneliti, diantaranya yaitu Indrayanti (2014) telah melakukan penelitian dengan membandingkan fungsi kernel *cosinus* dan kernel *gaussian* dengan menggunakan estimator *Nadaraya-Watson* pada data *butterfly diagram*

siklus aktivitas matahari ke-23 dan memperoleh hasil bahwa untuk data *butterfly diagram* siklus aktivitas matahari ke-23 fungsi kernel *cosinus* lebih baik dibandingkan dengan fungsi kernel *gaussian*. Nisa' (2016) juga telah membandingkan fungsi kernel *epanechnikov* dan kernel *triangle* pada rata-rata bulanan bilangan *sunspot* dan memperoleh hasil bahwa fungsi kernel *triangle* lebih baik digunakan untuk mengestimasi data rata-rata bulanan bilangan *sunspot* dibandingkan dengan fungsi kernel *epanechnikov*. Berdasarkan uraian di atas maka peneliti mengambil judul **“Estimasi Model Regresi Nonparametrik Menggunakan Estimator Nadaraya-Watson dengan Fungsi Kernel Epanechnikov”**.

1.2. Rumusan Masalah

Berdasarkan pada latar belakang di atas, maka rumusan masalah peneliti adalah bagaimana memperoleh estimasi parameter model regresi nonparametrik menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov* untuk data rata-rata harian bilangan *sunspot* pada tahun 2007.

1.3. Batasan Masalah

Agar mendapatkan hasil yang signifikan, maka dilakukan pembatasan masalah sebagai berikut :

1. Data yang digunakan adalah data rata-rata harian posisi *sunspot* pada lintang selatan.
2. Pemilihan nilai *bandwidth* dibatasi antara interval 0.1 sampai 2.0
3. Pemilihan *bandwidth* yang optimal berdasarkan *Generalized Cross Validation (GCV)*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah memperoleh penduga parameter model regresi nonparametrik menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov* untuk data rata-rata harian bilangan *sunspot* pada tahun 2007.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mampu mengetahui dan memahami regresi nonparametrik menggunakan estimator *Nadaraya-Watson* dengan fungsi kernel *Epanechnikov* menggunakan 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 estimator *Nadaraya-Watson*.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

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

$$Y_i = m(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

dengan ε_i adalah sisaan yang diasumsikan independen dengan nilai mean 0 dan variansi σ^2 .

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.

2.2 Regresi Nonparametrik

Analisis regresi nonparametrik merupakan metode statistika yang digunakan untuk melihat hubungan antara variabel respon dan variabel prediktor. Regresi nonparametrik tidak mensyaratkan bentuk sebaran parameter populasi sehingga dapat digunakan pada data yang memiliki sebaran normal atau tidak. Kurva regresi hanya diasumsikan dalam suatu fungsi yang berdimensi tak hingga dan merupakan fungsi pemulus (*smoothing*). Terdapat beberapa teknik *smoothing* dalam regresi nonparametrik antara lain *histogram*, *estimator spline*, *estimator kernel*, *deret fourier*, *deret orthogonal*, *k-NN*, dan *estimator wafelet*. Pendekatan nonparametrik tersebut memiliki keunggulannya masing-masing, misalnya estimator kernel yang memiliki bentuk lebih fleksibel dan lebih mudah disesuaikan dalam perhitungan matematisnya.

2.3 Estimator kernel

Estimator kernel diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga disebut estimator densitas kernel Rosenblatt-Parzen (Hardle W. , 1994). Estimator kernel merupakan suatu metode pendekatan terhadap fungsi densitas yang belum diketahui dengan menggunakan fungsi kernel. Penghalusan dengan pendekatan kernel dikenal sebagai pemulus kernel (*kernel smoother*) yang sangat bergantung pada fungsi kernel dan *bandwidth* (Saputra & Listyani, 2016).

Estimator kernel sangat umum digunakan untuk pendekatan nonparametrik. Beberapa kelebihan estimator kernel adalah fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara & Mulianah, 2007). Menurut Halim dan Bisono (2006:75) estimator kernel dibagi menjadi tiga macam, yaitu :

(1) Estimator *Nadaraya-Watson*

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)y_i}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)} \quad (2.2)$$

(2) Estimator *Priestley Chao*

$$\hat{m}(x_i) = \frac{1}{h} \sum_{i=1}^n (X_i - X_{i-1}) K\left(\frac{X_i-x}{h}\right) y_i \quad (2.3)$$

(3) Estimator *Gasser-Muller*

$$\hat{m}(x) = \frac{1}{h} \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} K\left(\frac{X_i-x}{h}\right) dx \quad (2.4)$$

$$\text{dimana } s_0 = 0, s_i = \frac{(X_i+X_{i+1})}{2}, i = 1,2,3, \dots, n-1, s_n = 1$$

Keterangan :

$\hat{m}(x)$ = fungsi taksiran regresi

y_i = variabel respon pada pengamatan ke-i

X_i = variabel prediktor pada pengamatan ke-i

x = nilai random variabel respon X atau nilai tertentu dari X

K = kernel

n = banyak pengamatan

h = *bandwidth*

2.4 Fungsi Kernel

Menurut Nisa'(2016:11) secara umum fungsi kernel didefinisikan sebagai berikut :

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \quad \text{untuk } -\infty < x < \infty, h > 0 \quad (2.5)$$

dengan K adalah fungsi kernel dan h adalah *bandwidth*. Fungsi kernel diatas harus memenuhi beberapa syarat, yaitu :

- (i) $K(x) \geq 0$, untuk semua x
- (ii) $\int_{-\infty}^{\infty} K(x)dx = 1$
- (iii) $\int_{-\infty}^{\infty} x^2 K(x)dx = \sigma^2 > 0$
- (iv) $\int_{-\infty}^{\infty} xK(x)dx = 0$

Terdapat beberapa jenis fungsi kernel yang umum digunakan untuk estimasi data, sebagai berikut.

- (i) Kernel Uniform : $K(x) = \frac{1}{2}I$; $|x| \leq 1$
- (ii) Kernel Triangle : $K(x) = (1 - |x|)I$; $|x| \leq 1$
- (iii) Kernel Epanechnikov : $K(x) = \frac{3}{4}(1 - x^2)I$; $|x| \leq 1$
- (iv) Kernel Quartik : $K(x) = \frac{15}{16}(1 - x^2)^2I$; $|x| \leq 1$
- (v) Kernel Triweight : $K(x) = \frac{35}{32}(1 - x^2)^3I$; $|x| \leq 1$
- (vi) Kernel Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}I$; $-\infty < x < \infty$
- (vii) Kernel Cosinus : $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right)I$; $|x| \leq 1$

dimana I adalah fungsi indikator, dengan

$$I(x) = \begin{cases} 1 & \text{jika } |x| \leq 1 \\ 0 & \text{jika } |x| > 1 \end{cases}, \text{ untuk } -\infty < x < \infty \quad (\text{Sudarno, 2011}).$$

Fungsi kernel yang akan digunakan dalam penelitian ini adalah fungsi kernel *Epanechnikov*. Menurut Hardle (1990) menyatakan bahwa kernel *Epanechnikov* merupakan kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding dengan kernel lainnya.

Salah satu metode untuk mengestimasi model regresi nonparametrik pada Persamaan (2.1) adalah regresi kernel. Regresi kernel merupakan teknik statistik

nonparametrik untuk menaksir nilai ekspektasi bersyarat dari suatu variabel acak. Nilai ekspektasi umumnya dinotasikan dengan $E(y|x)$. Regresi kernel bertujuan untuk mendapatkan hubungan nonlinier antara variabel x dan y . Ekspektasi bersyarat y terhadap x dinyatakan sebagai berikut :

$$E(y|x) = m(x) \text{ atau } \hat{y} = m(x) = \int_{-\infty}^{\infty} \frac{yf(x,y)}{f(x)} dy \quad (2.6)$$

dimana $f(x, y)$ = fungsi kepadatan bersama dari (x, y)

$$f(x) = \text{fungsi kepadatan marginal } x \quad (\text{Indrayanti, 2014}).$$

2.5 Estimator Nadaraya-Watson

Pada tahun 1964 Nadaraya dan Watson mengusulkan suatu estimator regresi kernel untuk memperkirakan m sebagai rata-rata tertimbang secara lokal dengan menggunakan kernel sebagai fungsi pembobotan (Saputra & Listyani, 2016). Pada fungsi kepadatan peluang bersama $f(x, y)$ yang perlu dilakukan pertama kali adalah melakukan estimasi pembilang dan penyebut secara terpisah dengan menggunakan estimator kernel, sehingga diperoleh (Apriani, 2015) :

$$f(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{Y_i - y}{h_y}\right) \quad (2.7)$$

Diperoleh hasil estimasi dari pembilang yang disajikan pada persamaan berikut :

$$\int y f(x, y) dy = \frac{1}{nh_x h_y} \int y \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{Y_i - y}{h_y}\right) dy \quad (2.8)$$

misalkan

$$\int y K\left(\frac{Y_i - y}{h_y}\right) dy = y_i$$

Sehingga dapat dituliskan :

$$\int y f(x, y) dy = \frac{1}{nh_x h_y} \sum K\left(\frac{X_i - x}{h_x}\right) y_i \quad (2.9)$$

Kemudian akan dihitung estimasi penyebut :

$$\int f(x, y) dy = \int \frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{Y_i - y}{h_y}\right) dy$$

misalkan

$$\int K\left(\frac{Y_i - y}{h_y}\right) dy = 1$$

Maka diperoleh,

$$\int f(x, y) dy = \frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) \quad (2.10)$$

Jadi, estimasi kernel *Nadaraya-Watson* untuk $\hat{m}(x)$ dari fungsi regresi yaitu (D'Addona C. M., 2017) :

$$\begin{aligned} \hat{m}(x) &= \frac{\frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) y_i}{\frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right)} \\ &= \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right) y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_x}\right)} \end{aligned} \quad (2.11)$$

Dimana $h_x = h$, sehingga dapat dituliskan dalam Persamaan (2.12) :

$$\begin{aligned} \hat{m}(x) &= \frac{\frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) y_i}{\frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \\ &= \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \end{aligned} \quad (2.12)$$

dengan :

x = nilai random variabel X atau nilai tertentu dari variabel X

h = *bandwidth*

X_i = nilai variabel prediktor pada pengamatan ke- i

y_i = nilai variabel respon pada pengamatan ke- i

2.6 Pemilihan *Bandwidth* Optimum

Menurut Hardle (1991), *bandwidth* (h) adalah parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Jika *bandwidth* yang dipilih terlalu kecil maka akan menghasilkan estimasi kurva kurang mulus (*under-smoothing*), sebaliknya jika *bandwidth* terlalu besar maka akan menghasilkan estimasi kurva yang sangat mulus (*over-smoothing*) yang tidak sesuai dengan pola sebaran 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 nilai error yang minimum. Salah satu

metode untuk mendapatkan *bandwidth* optimum diperoleh dengan menggunakan kriteria *Generalized Cross Validation* (GCV), yang didefinisikan sebagai berikut :

$$GCV = \frac{MSE}{\left(\frac{1}{n} \text{tr}(I-H(h))\right)^2} \quad (2.13)$$

dengan :

- n = banyaknya data
- I = matriks identitas
- h = *bandwidth*
- X = matriks data
- $H(h) = X(X'X + nhI)^{-1}X'$
- $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - m_h(x_i))^2$

Menurut Komang dan Gusti (2012), kebaikan suatu estimator dapat dilihat dari tingkat kesalahannya. Terdapat beberapa kriteria untuk menentukan estimator terbaik dalam model regresi nonparametrik, diantaranya :

a. *Means Square Error*

Untuk mengukur error biasanya digunakan *Mean Square Error*. Estimator terbaik dipilih berdasarkan nilai MSE terkecil. *Mean Square Error* adalah rata-rata dari kuadrat kesalahan.

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.14)$$

b. *Root Mean Square Error*

$$RMSE = \sqrt{MSE} \quad (2.15)$$

c. *Mean Absolute Deviation*

$$MAD = \frac{1}{n} \sum_{i=1}^n |e_i| = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.16)$$

Keterangan :

- n = banyaknya data
- y_i = nilai variabel respon
- \hat{y}_i = nilai prediksi dari variabel y_i

2.7 Matahari

Matahari merupakan salah satu benda terpenting yang berada di pusat tata surya. Menurut Kaufmann (1978), matahari adalah suatu bola gas yang berukuran sangat besar dengan ukuran diameter mencapai 1.400 kilometer. Matahari disebut juga bintang yang berada di pusat tata surya dan merupakan sumber energi di sekitar tata surya. Matahari disebut bintang karena dapat memancarkan cahaya sendiri. Plasma dalam matahari sangatlah panas, dengan suhu permukaan sekitar 5800 derajat Kelvin dan suhu tertinggi mencapai 15 juta derajat Kelvin pada intinya. Cahaya yang kita terima di bumi paling banyak datang dari permukaan matahari. Matahari mempunyai khatulistiwa dan kutub karena gerak rotasinya. Garis tengah ekuatornya 864.000 mil, sedangkan garis tengah antar kutubnya 43 mil lebih pendek. Matahari merupakan anggota tata surya yang paling besar, karena 98% massa tata surya terkumpul pada matahari. Matahari merupakan bintang terdekat dengan bumi dengan jarak rata-rata 149.680.000 (93.026.724 mil).

Beberapa lapisan matahari antara lain adalah lapisan inti (core), radiative, konvektif, fotosfer, kromosfer, dan lapisan corona. Fenomena tersebut muncul akibat berlangsungnya transformasi energi dari pusat matahari ke arah luar yang dilakukan secara radiasi maupun konveksi. Diantara aktivitas matahari tersebut yang sering diamati para ahli fisika adalah fenomena kemunculan sunspot. Fenomena sunspot ini dapat mengakibatkan flare yaitu ledakan matahari (Indrayanti, 2014). Sumber energi dari semua bentuk energi yang dipancarkan matahari berasal dari pusat matahari. Sinar matahari merupakan sumber kehidupan bagi makhluk di bumi terutama tumbuhan sebagai produsen di bumi. Matahari terdiri dari gas panas, untuk terus bersinar merubah hydrogen menjadi helium melalui reaksi nuklir. Selain itu, pada atmosfer matahari terjadi berbagai fenomena. Fenomena-fenomena ini disebut sebagai aktivitas matahari. Aktivitas-aktivitas matahari yang dapat diamati dari bumi terjadi pada lapisan *fotosfer*. Berbagai aktivitas matahari diantaranya *sunspot*, *granula*, *faculae*, *flare*, *filamen*, dan *prominensa*. *Sunspot* adalah daerah di lapisan fotosfer matahari, yang memiliki suhu yang lebih rendah dari pada sekelilingnya. *Granula* adalah hasil

dari udara yang muncul dari interior matahari yang menyebar setelah mencapai permukaan. *Faculae* adalah gelombang *white-light* sebagai titik terang dekat kutub-kutub utara dan selatan. *Flare* adalah kilatan cahaya energi tinggi. *Filamen* adalah awan gas terionisasi di atas permukaan matahari yang terhempas diantara daerah magnet dengan polaritas berlawanan. *Prominensa* adalah bagian matahari yang menyerupai lidah api yang sangat besar dan terang yang muncul keluar dari bagian permukaan serta sering kali berbentuk loop (putaran).

2.8 Sunspot Matahari

Sunspot matahari atau yang lebih dikenal sebagai bintik matahari adalah fenomena yang terjadi di fotosfer matahari dimana ada suatu daerah yang berwarna lebih gelap dibanding daerah disekitarnya. Pada permukaan matahari terdapat bercak-bercak gelap sehingga fotosfer kelihatan tidak terlalu mulus. Daerah-daerah ini dinamakan daerah *sunspot* (bintik matahari). Bintik matahari bukanlah daerah yang tidak memancarkan cahaya, hanya saja temperatur daerah ini lebih rendah dibandingkan daerah sekitarnya (3.000-4.500 derajat Celcius) sehingga kelihatan gelap (Admiranto, 2009).

Bintik matahari di fotosfer adalah daerah dengan medan magnet yang kuat yang lebih dingin, dan dengan demikian lebih gelap, dibandingkan dengan wilayah sekitarnya. Nomor sunspot berfluktuasi setiap 11 tahun sebagai bagian dari siklus aktivitas magnetik matahari (Nguyen, 2018). Bintik matahari terdiri dari dua bagian, yaitu bagian pusat yang paling gelap berwarna umbra, dan dikelilingi bagian yang lebih terang yang disebut dengan penumbra. Bintik matahari sering didapat berpasangan. Bintik-bintik ini jarang yang bias bertahan cukup lama, ada yang umumnya kurang dari satu hari, ada juga yang umurnya mencapai beberapa bulan. Seorang astronomi amatir dari Jerman bernama Heinrich Schwabe mendapati bahwa jumlah bintik matahari berubah-ubah secara periodik. Ia mengamati bahwa jumlah bintik matahari berubah-ubah dengan periode rata-rata 10,5 tahun (Admiranto, 2009). Analisis perioditas menunjukkan bahwa bilangan *sunspot* lebih baik digunakan sebagai dasar prakiraan aktivitas matahari. Perioditas utamanya (sekitar 11 tahun) relatif konstan sehingga dapat

digunakan untuk perkiraan jangka panjang beberapa siklus ke depan (Djamaluddin, 2005). Dalam pengamatan *sunspot*, indeks yang paling penting adalah *sunspot number* atau bilangan *sunspot*. Bilangan *sunspot* merupakan parameter yang menyatakan tingkat aktivitas matahari. Aktivitas matahari yang meningkat dan memiliki banyak bintik akan memicu terjadinya ledakan matahari (*flare*).