

**ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN  
METODE NADARAYA-WATSON DENGAN FUNGSI KERNEL  
GAUSSIAN**

(Skripsi)

Oleh

**MEILINDA**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2019**

## **ABSTRACT**

### **ESTIMATION OF NONPARAMETRIC REGRESSION CURVE USING NADARAYA-WATSON METHOD WITH GAUSSIAN KERNEL FUNCTION**

**By**

**MEILINDA**

The purpose of this research is to know whether the estimation of nonparametric curve using Nadaraya-Watson method with the Gaussian Kernel Function depends on the error distribution. Simulation data is obtained from linear function with different error distribution. The results of this research shows that the distribution of error does not affect the estimation of regression curve using Nadaraya-Watson Method. Nadaraya-Watson method with kernel Gaussian function can be used to estimate the regression curve on linear function that has normally distributed error and non normally distributed error.

**Keywords:** Regression Nonparametric, Nadaraya-Watson Method, Error

## **ABSTRAK**

### **ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN METODE NADARAYA-WATSON DENGAN FUNGSI KERNEL GAUSSIAN**

**Oleh**

**MEILINDA**

Penelitian ini bertujuan untuk mengetahui apakah estimasi kurva regresi nonparametrik menggunakan metode Nadaraya-Watson dengan fungsi Kernel Gaussian bergantung pada distribusi galat. Data simulasi diperoleh dari fungsi linear dengan distribusi galat yang berbeda. Hasil penelitian menunjukkan bahwa distribusi galat tidak mempengaruhi estimasi kurva regresi menggunakan metode Nadaraya Watson. Metode Nadaraya-Watson dengan fungsi Kernel Gaussian dapat digunakan untuk mengestimasi kurva regresi pada fungsi linear yang memiliki galat berdistribusi normal dan galat berdistribusi tidak normal.

**Kata Kunci:** Regresi Nonparametrik, Metode Nadaraya-Watson, Galat

**ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN  
METODE NADARAYA-WATSON DENGAN FUNGSI KERNEL  
GAUSSIAN**

**Oleh**

**MEILINDA**

**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 KURVA REGRESI  
NONPARAMETRIK MENGGUNAKAN  
METODE NADARAYA-WATSON DENGAN  
FUNGSI KERNEL GAUSSIAN**

Nama Mahasiswa : Meilinda


Nomor Pokok Mahasiswa : 1517031007


Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

**MENYETUJUI**

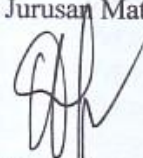
1. Komisi Pembimbing

  
**Ir. Netti Herawati, M.Sc., Ph.D.**  
NIP. 19650125 199003 2 001

  
**Dr. Notiragayu, S.Si., M.Si.**  
NIP. 19731109 200012 2 001

2. Mengetahui

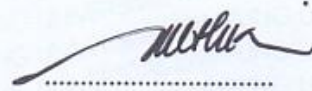
Ketua Jurusan Matematika

  
**Prof. Dra. Wamiliana, M.A., Ph.D.**  
NIP. 19631103 198902 2 001

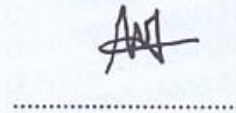
MENGESAHKAN

1. Tim Penguji

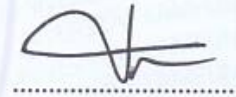
Ketua : Ir. Netti Herawati, M.Sc., Ph.D.



Sekretaris : Dr. Notiragayu, S.Si., M.Si.



Penguji  
Bukan Pembimbing : Drs. Nusyirwan, M.Si.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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

Tanggal Lulus Ujian Skripsi: 14 Agustus 2019

## SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Meilinda

Nomor Pokok Mahasiswa : 1517031007

Judul : **Estimasi Kurva Regresi Nonparametrik  
Menggunakan Metode Nadaraya-Watson dengan  
Fungsi Kernel Gaussian**

Jurusan : Matematika

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, Agustus 2019

Penulis



  
**Meilinda**  
NPM. 1517031007

## **RIWAYAT HIDUP**

Penulis bernama lengkap Meilinda, anak pertama dari dua bersaudara yang dilahirkan di Gayau Sakti pada tanggal 4 Mei 1997 oleh pasangan Bapak Sutiyo dan Ibu Wartini.

Penulis menyelesaikan pendidikan Sekolah dasar di SD Negeri 2 Gayau Sakti pada tahun 2009. Sekolah menengah pertama di SMP Negeri 1 Way Pengubuan pada tahun 2012. Sekolah menengah atas di SMA Negeri 1 Seputih Agung 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 Tengah dan sebagai bentuk pengabdian kepada masyarakat, penulis telah melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Sinar Mulyo, Kecamatan Pulau Panggung, Kabupaten Tanggamus.



## KATA INSPIRASI

*“Man Jadda Wajada”*

*(siapa yang bersungguh-sungguh akan berhasil)*

*“Hai orang-orang yang beriman, jadikanlah sabar dan sholat sebagai penolongmu, sesungguhnya Allah bersama orang-orang yang sabar”*

*(Q.S. Al-Baqarah:153)*

*“Boleh jadi kamu membenci sesuatu, padahal itu amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal itu amat buruk bagimu. Allah yang paling mengetahui, sedangkan kamu tidak mengetahui”*

*(Q.S. Al-Baqarah:216)*

## PERSEMBAHAN

Alhamdulillahirobbil'alamin,

Puji dan syukur tiada hentinya terpanjatkan kepada Allah SWT

Ku persembahkan karya kecil ini untuk:

Ibu dan Bapak, serta Adikku Tercinta yang selalu tulus mendoakan setiap waktu,  
membimbing, dan selalu memberikan semangat dan motivasi untuk keberhasilan

penulis

Dosen-dosen Pembimbing dan Pembahas yang sangat berjasa dan selalu  
memberikan motivasi kepada penulis.

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

Almamater kebanggaan, Universitas Lampung

## SANWACANA

Puji dan syukur penulis panjatkan kepada Allah SWT atas izin serta ridho-Nya penulis dapat menyelesaikan skripsi yang berjudul “Estimasi Kurva Regresi Nonparametrik Menggunakan Metode Nadaraya-Watson dengan Fungsi Kernel Gaussian”. 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 sekaligus Dosen Pembimbing Akademik, yang senantiasa selalu membimbing dan memberikan arahan, ide, kritik dan saran serta semangat kepada penulis selama proses pembuatan skripsi ini.
2. Ibu Dr. Notiragayu, S.Si., M.Si., selaku Dosen Pembimbing II, yang telah memberikan bimbingan serta saran yang membantu penulis dal menyelesaikan skripsi ini.
3. Bapak Drs. Nusyirwan, M.Si., selaku Dosen Pembahas, terima kasih atas kesediaannya untuk membahas, memberikan saran dan kritik yang membangun dalam penyelesaian skripsi ini.

4. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Drs. Suratman, M.Sc., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh Dosen, Staf, dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. 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.
8. Sahabat-sahabat tercinta: Trisna, Yunita, May, Reni, Amirah, Aulia Safitri, Purwanti, Rosita, Azam, Sindi, Riswanti, dan El yang selalu menemani hari-hari penulis selama menjalani masa perkuliahan dan memberikan semangat untuk menyelesaikan skripsi ini.
9. Teman-temanku Matematika 2015, terimakasih telah memberikan warna dan keceriaan kepada penulis selama menjadi mahasiswi.
10. Almamater tercinta Universitas Lampung.
11. Seluruh pihak yang telah membantu yang tidak dapat disebutkan satu persatu.

Bandar Lampung, Agustus 2019

Penulis,

**Meilinda**

## DAFTAR ISI

	Halaman
<b>DAFTAR GAMBAR</b> .....	iii
<b>DAFTAR TABEL</b> .....	v
<b>I. PENDAHULUAN</b> .....	1
1.1 Latar Belakang dan Masalah.....	1
1.2 Tujuan Penelitian .....	3
1.3 Manfaat Penelitian .....	3
<b>II. TINJAUAN PUSTAKA</b> .....	4
2.1 Analisis Regresi .....	4
2.2 Regresi Nonparametrik .....	5
2.3 Estimator Kernel .....	6
2.4 Metode Nadaraya-Watson.....	7
2.5 Fungsi Kernel.....	8
2.6. Metode Nadaraya-Watson dengan Fungsi Kernel Gaussian.....	10
2.7 <i>Bandwidth</i> Optimal .....	11
2.8 Kelayakan Model .....	12
<b>III. METODOLOGI PENELITIAN</b> .....	14
3.1 Waktu dan Tempat Penelitian .....	14
3.2 Data Penelitian .....	14
3.2 Metode Penelitian .....	14
<b>IV. HASIL DAN PEMBAHASAN</b> .....	16
4.1 Diagram Pencar Data Pengamatan.....	16
4.2 Estimasi Kurva Regresi Nonparametrik Fungsi $y=1+12x$ pada Distribusi Galat Berbeda.....	18
4.2.1 Penentuan <i>Bandwidth</i> optimal Metode Nadaraya-Watson pada fungsi $y=1+12x$ dengan galat $e \sim N(0,1)$ .....	18
4.2.2 Penentuan <i>Bandwidth</i> optimal Metode Nadaraya-Watson pada fungsi $y=1+12x$ dengan galat $e_2 \sim N(0,0.3)$ .....	21
4.2.3 Penentuan <i>Bandwidth</i> Optimal Metode Nadaraya-Watson pada fungsi $y=1+12x$ dengan galat $e_3 \sim 0.9 N(0,1) + 10\%$ pencilan dari $N(5,1)$ .....	24

4.2.4	Penentuan <i>Bandwidth</i> optimal Metode Nadaraya-Watson pada fungsi $y=1+12x$ dengan galat $e_4 \sim U(-2,2)$ .....	26
4.3	Perbandingan Ukuran Kebaikan Model Menggunakan Metode Nadaraya-Watson pada Fungsi $y=1+12x$ dengan Beberapa Distribusi Galat .....	29
<b>V.</b>	<b>KESIMPULAN</b> .....	31
	<b>DAFTAR PUSTAKA</b> .....	32
	<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar	Halaman
1. Grafik fungsi kernel .....	9
2. Diagram pencar fungsi $y = 1 + 12x + e_1 \sim N(0,1)$ .....	16
3. Diagram pencar fungsi $y = 1 + 12x + e_2 \sim N(0,0.3)$ .....	17
4. Diagram pencar fungsi $y = 1 + 12x + e_3 \sim 0.9N(0,1) + 10\%$ pencilan dari $N(5,1)$ .....	17
5. Diagram pencar fungsi $y = 1 + 12x + e_4 \sim U(-2,2)$ .....	18
6. Kurva Estimasi Nadaraya-Watson pada fungsi $y = 1 + 12x + e_1 \sim N(0,1)$ dengan $h$ optimal = 0,1085 .....	20
7. Kurva Estimasi Nadaraya-Watson pada fungsi $y = 1 + 12x + e_1 \sim N(0,1)$ dengan $h = 0,01$ .....	20
8. Kurva Estimasi Nadaraya-Watson pada fungsi $y = 1 + 12x + e_1 \sim N(0,1)$ dengan $h = 0,515$ .....	20
9. Kurva Estimasi Nadaraya-Watson fungsi $y = 1 + 12x + e_2 \sim N(0,0.3)$ dengan $h$ optimal = 0,0818 .....	22
10. Kurva Estimasi Nadaraya-Watson fungsi $y = 1 + 12x + e_2 \sim N(0,0.3)$ dengan $h = 0,001$ .....	23
11. Kurva Estimasi Nadaraya-Watson fungsi $y = 1 + 12x + e_2 \sim N(0,0.3)$ dengan $h = 0,5818$ .....	23
12. Kurva Estimasi Nadaraya-Watson fungsi $y = 1 + 12x + e_3 \sim 0.9N(0,1)$ + 10% pencilan dari $N(5,1)$ dengan $h$ optimal = 0,2024 .....	25
13. Kurva Estimasi Nadaraya-Watson fungsi $y = 1 + 12x + e_3 \sim 0.9N(0,1)$ + 10% pencilan dari $N(5,1)$ dengan $h = 0,001$ .....	25

14. Kurva Estimasi Nadaraya-Watson fungsi $y = 1+12x + e_3 \sim 0.9N(0,1)$ + 10% pencilan dari $N(5,1)$ dengan $h = 0,5324$ .....	26
15. Kurva Estimasi Nadaraya-Watson fungsi $y = 1+12x + e_4 \sim U(-2,2)$ dengan $h$ optimal= $0,2745$ .....	27
16. Kurva Estimasi Nadaraya-Watson fungsi $y = 1+12x + e_4 \sim U(-2,2)$ dengan $h = 0,001$ .....	28
17. Kurva Estimasi Nadaraya-Watson fungsi $y = 1+12x + e_4 \sim U(-2,2)$ dengan $h = 0,5745$ .....	28



## DAFTAR TABEL

Tabel	Halaman
1. Nilai <i>Generalized Cross Validation</i> (GCV) Metode Nadaraya-Watson pada Fungsi $y=1+12x$ dengan galat $e_1 \sim N(0,1)$ .....	19
2. Nilai <i>Generalized Cross Validation</i> (GCV) Metode Nadaraya-Watson pada Fungsi $y = 1 + 12x$ dengan galat $e_2 \sim N(0,0.3)$ .....	21
3. Nilai <i>Generalized Cross Validation</i> (GCV) Metode Nadaraya-Watson pada Fungsi $y=1+12x$ dengan galat $e_3 \sim 0.9 N(0,1) + 10\%$ pencilan dari $N(5,1)$ .....	24
4. Nilai <i>Generalized Cross Validation</i> (GCV) metode Nadaraya-Watson pada fungsi $y=1+12x$ dengan galat $e_4 \sim U(-2,2)$ .....	27
5. Daftar nilai ukuran kebaikan model menggunakan metode Nadaraya-Watson pada fungsi $y = 1+12x$ dengan beberapa distribusi galat.....	29

## I. PENDAHULUAN

### 1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu teknik statistika yang digunakan untuk menggambarkan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor. Pada regresi harus ada variabel yang ditentukan dan variabel yang menentukan atau dengan kata lain adanya ketergantungan antara variabel yang satu dengan yang lainnya (Hosmer and Lemeshow, 2000). Cara sederhana untuk mengetahui informasi mengenai hubungan antara variabel respon dan variabel prediktor dapat dilihat melalui bentuk pola hubungan pada diagram pencar. Dengan mengetahui pola hubungan yang terbentuk, dapat ditentukan pendekatan yang sesuai untuk mengestimasi fungsi regresi.

Menurut Netter, *et al.*, (1997), untuk mengestimasi fungsi regresi ada dua pendekatan yang dapat digunakan yaitu secara parametrik dan nonparametrik. Jika bentuk kurva regresi diketahui, maka untuk menduga fungsi regresi dapat menggunakan pendekatan parametrik. Untuk dapat menggunakan metode ini, data harus memenuhi asumsi normalitas. Jika asumsi tidak terpenuhi, maka digunakan analisis regresi nonparametrik.

Regresi nonparametrik dapat digunakan untuk memodelkan data yang berbentuk apa saja, baik linear maupun nonlinear dikarenakan tidak adanya asumsi yang harus dipenuhi. Estimasi fungsi regresi nonparametrik menggunakan teknik *smoothing*. Terdapat beberapa teknik *smoothing* dalam model regresi nonparametrik antara lain histogram, estimator kernel, estimator *spline*, deret *fourier*, dan *wavelet* (Eubank, 1998).

Estimator kernel adalah pengembangan dari estimator histogram. Estimator ini merupakan estimator linier yang mirip dengan estimator regresi nonparametrik yang lain, perbedaannya hanya karena estimator kernel lebih khusus dalam penggunaan *bandwidth*. Kelebihan dari estimator kernel adalah memiliki kemampuan yang baik dalam memodelkan data yang tidak mempunyai pola tertentu (Hardle, 1994). Pada teknik pemulusan menggunakan estimator kernel, metode yang sering digunakan yaitu Nadaraya-Watson. Metode kernel ini membutuhkan fungsi kernel dan *bandwidth* yang optimal dalam pengoperasiannya karena pemilihan *bandwidth* merupakan langkah penting yang harus dilakukan pada teknik pemulusan. Salah satu metode untuk mencari *bandwidth* yang optimal adalah menggunakan kriteria *Generalized Cross Validation* (GCV).

Pada penelitian ini akan dilakukan estimasi kurva regresi nonparametrik menggunakan metode Nadaraya-Watson dengan fungsi Kernel Gaussian pada data bangkitan dengan distribusi galat berbeda, serta pemilihan *bandwidth* menggunakan metode *Generalized Cross Validation* (GCV).

## 1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menentukan *bandwidth* optimal pada metode Nadaraya-Watson dengan fungsi Kernel Gaussian menggunakan *Generalized Cross Validation* (GCV) pada data bangkitan dengan distribusi galat berbeda.
2. Untuk mengetahui apakah estimasi kurva regresi nonparametrik menggunakan metode Nadaraya-Watson dengan fungsi Kernel Gaussian bergantung pada distribusi galat.

## 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 Gaussian.
2. Menambah referensi tentang regresi nonparametrik menggunakan metode Nadaraya-Watson dengan fungsi Kernel Gaussian.

## **II. TINJAUAN PUSTAKA**

### **2.1 Analisis Regresi**

Analisis regresi merupakan salah satu teknik statistika yang sering digunakan untuk mengetahui hubungan kausal antara variabel respon dan variabel prediktor. Untuk data yang satu variabel prediktor dan satu variabel respon, analisis regresi disebut juga dengan analisis regresi linear sederhana. Sedangkan untuk data yang memiliki lebih dari satu variabel prediktor dan satu variabel respon disebut dengan analisis regresi linear berganda. Jadi secara umum analisis regresi adalah metode analisis data yang menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor (Hosmer and Lemeshow, 2000).

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

Statistik nonparametrik dapat digunakan pada data yang memiliki distribusi normal ataupun tidak. Istilah nonparametrik pertama kali diperkenalkan oleh Wolfowitz pada tahun 1942. Pendekatan nonparametrik merupakan pendekatan regresi yang sesuai untuk pola data yang tidak diketahui bentuknya, atau tidak terdapat informasi masa lalu tentang pola data (Lestari dan Budiantara, 2010). Menurut Hardle (1990), pendekatan nonparametrik merupakan pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu. Kurva regresi yang sesuai dengan pendekatan nonparametrik diwakili oleh model yang disebut dengan model regresi nonparametrik.

Regresi nonparametrik merupakan pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi. Regresi nonparametrik dapat digunakan untuk memodelkan data yang berbentuk apa saja, baik linear maupun nonlinear dikarenakan tidak adanya asumsi yang harus dipenuhi. Model regresi nonparametrik secara matematis dapat ditulis:

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

dengan  $y_i$  adalah variabel respon pada pengamatan ke- $i$ ,  $\varepsilon_i$  merupakan galat/sisaan/kesalahan ke- $i$  yang diasumsikan menyebar normal dengan mean sama dengan 0 dan ragam tetap, serta  $m(x_i)$  merupakan fungsi regresi atau kurva regresi (Suparti, dkk., 2018).

Dalam regresi nonparametrik, 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*). Estimasi fungsi  $m(x_i)$  dilakukan berdasarkan data pengamatan dengan menggunakan teknik *smoothing* tertentu. Ada beberapa teknik *smoothing* yang dapat digunakan antara lain estimator histogram, kernel, penduga spline, deret fourier, dan wavelet (Eubank, 1998).

### 2.3 Estimator Kernel

Estimator 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*. Pemilihan *bandwidth* akan memengaruhi hasil *smoothing* kernel. Nilai *bandwidth* yang semakin kecil akan menyebabkan bentuk kurva semakin kasar dan sebaliknya semakin besar nilai *bandwidth* akan menyebabkan kurva semakin mulus dan tidak sesuai dengan pola data. Estimator kernel memiliki beberapa keuntungan diantaranya:

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

Menurut Halim dan Bisono (2006), estimator kernel dibagi menjadi tiga macam, yaitu: Estimator Nadaraya-Watson, Priestley-Chao Estimate, dan Gasser-Muller Estimate. Metode yang sering dipakai adalah metode Nadaraya-Watson karena perhitungannya lebih sederhana dibandingkan estimator kernel lainnya.

## 2.4 Metode Nadaraya-Watson

Menurut Hardle (1991), jika terdapat  $n$  data pengamatan  $\{(X_i, Y_i)\}_{i=1}^n$  yang memenuhi persamaan (2.1) dimana  $X_i \in \mathbb{R}$  dan  $Y_i \in \mathbb{R}$ , maka penduga  $m(x)$  adalah :

$$E(Y|X = x) = \hat{m}(x) = \int \frac{yf(x,y)}{f(x=x)} dy \quad (2.2)$$

Penyebut diduga dengan menggunakan penduga densitas kernel

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

Fungsi densitas peluang bersama diduga dengan perkalian kernel, yaitu:

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

Sehingga, pembilang dari penduga Nadaraya-Watson menjadi:

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

Bentuk penduga Nadaraya-Watson dapat ditulis:

$$\hat{m}(x_i) = \frac{\frac{1}{n} \sum_{j=1}^n K_h(x_i - X_j) Y_j}{\frac{1}{n} \sum_{j=1}^n K_h(x_i - X_j)}$$



$$\hat{m}(x_i) = \frac{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right) Y_j}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right)}$$

$$\hat{m}(x_i) = \frac{\sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right) Y_j}{\sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right)} \quad (2.3)$$

$$\hat{m}(x_i) = \sum_{j=1}^n W_{ij}(x_i) Y_j$$

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

$$W_{ij}(x_i) = \frac{K\left(\frac{x_i - X_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right)} \quad (2.4)$$

Persamaan (2.3) ditemukan oleh Nadaraya dan Watson (1964), sehingga disebut estimator Nadaraya-Watson. Sedangkan, Matriks W pada persamaan (2.4) disebut *Hat Matrix* berukuran  $n \times n$  dari penduga  $m(x_i)$  pada metode Nadaraya-Watson.

## 2.5 Fungsi Kernel

Menurut Wand and Jones, (1995) secara umum kernel K dengan parameter pemulus *bandwidth* (h) didefinisikan sebagai :

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

Serta memenuhi :

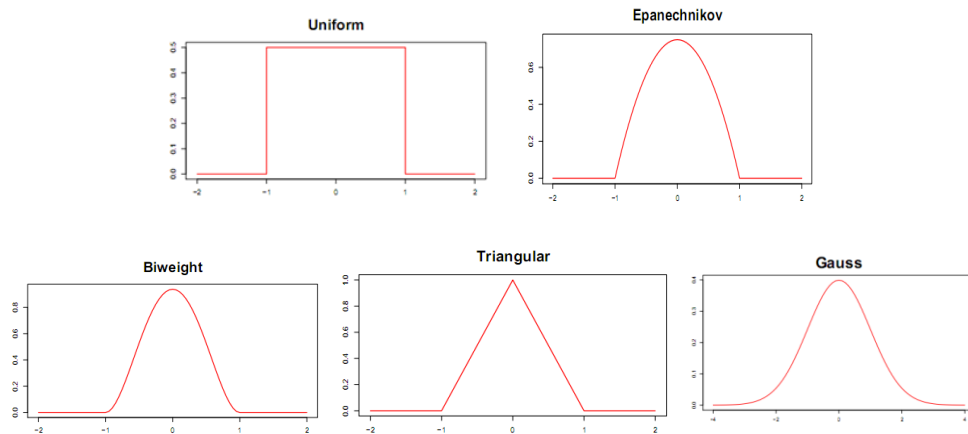
- (i)  $K(x) \geq 0$ , untuk semua x
- (ii)  $\int_{-\infty}^{\infty} K(x) dx = 1$
- (iii)  $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$

$$(iv) \int_{-\infty}^{\infty} xK(x)dx = 0$$

Beberapa jenis fungsi kernel antara lain:

1. Kernel Uniform :  $K(x) = \frac{1}{2}$  ;  $|x| \leq 1, 0$  selainnya
2. Kernel Triangle :  $K(x) = (1-|x|)$  ;  $|x| \leq 1, 0$  selainnya
3. Kernel Epanechnikov :  $K(x) = \frac{3}{4}(1-x^2)$  ;  $|x| \leq 1, 0$  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 Cosinus :  $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right)$  ;  $|x| \leq 1, 0$  selainnya
7. Kernel Gaussian :  $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$  ;  $-\infty < x < \infty$

Berikut bentuk grafik dari fungsi kernel diatas,



Gambar 1. Grafik fungsi kernel

## 2.6 Metode Nadaraya-Watson dengan Fungsi Kernel Gaussian

Estimator Nadaraya-Watson pada persamaan (2.3) didefinisikan sebagai berikut:

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

Dengan  $\hat{m}(x_i)$  adalah penduga  $m(x_i)$ ,  $K(\cdot)$  adalah fungsi kernel, dan  $h$  adalah parameter pemulus (*bandwidth*). Pada penelitian ini fungsi kernel  $K$  yang digunakan adalah fungsi Kernel Gaussian yang didefinisikan sebagai berikut:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}; -\infty < x < \infty \quad (2.6)$$

Maka akan diperoleh,

$$\begin{aligned} \hat{m}(x_i) &= \frac{\sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right) Y_j}{\sum_{j=1}^n K\left(\frac{x_i - X_j}{h}\right)} \\ \hat{m}(x_i) &= \frac{\sum_{j=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - X_j}{h}\right)^2} Y_j}{\sum_{j=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - X_j}{h}\right)^2}} \end{aligned} \quad (2.7)$$

$$\hat{m}(x_i) = \sum_{j=1}^n W_{ij}(x_i) Y_j$$

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

$$W_{ij}(x_i) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - X_j}{h}\right)^2}}{\sum_{j=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - X_j}{h}\right)^2}} \quad (2.8)$$

Persamaan (2.7) merupakan penduga Nadaraya-Watson dengan menggunakan Kernel Gaussian. Sedangkan, Matriks  $W$  pada persamaan (2.8) disebut juga dengan *Hat Matrix* berukuran  $n \times n$  dari penduga  $m(x_i)$  pada metode Nadaraya-Watson.

## 2.7 Bandwidth Optimal

*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, sebaliknya *bandwidth* yang terlalu besar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data, maka dari itu perlu dipilih *bandwidth* yang optimal (Hardle, 1991). Dalam prakteknya, pemilihan *bandwidth* biasanya dilakukan dengan cara trial and error. Silverman (1986), memberikan lebar jendela (*bandwidth*) dengan rumus sebagai berikut:

$$h = 0.9 \min \left\{ S, \frac{IQR}{1.34} \right\} n^{-1/5} \quad (2.9)$$

dengan,  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ , IQR adalah Inter Quartil Range (jangkauan antar kuartil= $Q_3 - Q_1$ ), dan  $n$  adalah banyaknya data.

Untuk menghasilkan kurva optimal diperlukan *bandwidth* yang optimal. Metode untuk mendapatkan  $h$  optimal dapat diperoleh dengan menggunakan *Generalized Cross Validation* (GCV) yang didefinisikan sebagai berikut:

$$GCV(h) = \frac{MSE}{\left(1 - \frac{\text{trace}(W)}{N}\right)^2} \quad (2.10)$$

dengan  $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{m}(x_i))^2$  dan  $W$  adalah hat matriks berukuran  $n \times n$  yang memenuhi  $[\hat{m}(x_1), \hat{m}(x_2), \dots, \hat{m}(x_n)]^t = \mathbf{WY}$  (Persamaan 2.8). Nilai *bandwidth* optimal akan diperoleh jika menghasilkan nilai *Generalized Cross Validation* minimal (Craven and Wahba, 1979).

## 2.8 Kelayakan Model

Kebaikan suatu penduga dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahan suatu pendugaan maka semakin baik estimasinya. Menurut Chatterjee (2006), kriteria untuk menentukan estimator terbaik dalam model regresi antara lain dengan melihat nilai *Mean Square Error* (MSE) dan nilai koefisien determinasi *R-Square* ( $R^2$ ). MSE didefinisikan sebagai berikut :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.11)$$

Sedangkan koefisien determinasi didefinisikan sebagai berikut :

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.12)$$

$y_i$  adalah data variabel respon ke- $i$ ,  $\bar{y}$  adalah mean data variabel respon, sedangkan  $\hat{y}_i$  adalah nilai hasil estimasi variabel respon ke- $i$ . *Sum of Square Regression* (SSR) adalah jumlah kuadrat simpangan hasil dugaan terhadap rata-rata variabel respon, sedangkan *Sum of Square Total* (SST) adalah jumlah kuadrat simpangan variabel respon. SSR berfungsi untuk mengukur kualitas variabel prediktor sebagai prediktor variabel respon. Sehingga, koefisien determinasi dapat diartikan sebagai proporsi keragaman total variabel respon yang diukur oleh variabel prediktor. Nilai koefisien determinasi adalah diantara nol dan satu. Nilai  $R^2$  yang kecil berarti kemampuan variabel-variabel prediktor dalam menjelaskan variasi variabel

respon amat terbatas. Nilai yang mendekati satu berarti variabel-variabel prediktor memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variansi variabel respon.

### III. METODOLOGI PENELITIAN

#### 3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 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 bangkitan menggunakan aplikasi R dengan variabel bebas ( $x$ ) berdistribusi uniform (1,3) sebanyak 30 data dan variabel tak bebas ( $y$ ) diperoleh dari fungsi  $y_i = 1 + 12x + e_i$ , dengan galat

1.  $e_1 \sim N(0,1)$ , galat normal dengan  $\mu = 0$  dan  $\sigma^2 = 1$ .
2.  $e_2 \sim N(0,0.3)$ , galat normal dengan  $\mu = 0$  dan  $\sigma^2 = 0,3$ .
3.  $e_3 \sim 0.9N(0,1) + 10\%$  pencilan dari  $N(5,1)$ .
4.  $e_4 \sim U(-2,2)$ .

#### 3.3 Metode Penelitian

Penelitian ini menggunakan metode estimator Nadaraya-Watson dengan fungsi Kernel Gaussian. Penentuan *bandwith* ( $h$ ) dilakukan dengan melihat *Generalized*

*Cross Validation* (GCV) minimum. Data diolah menggunakan Aplikasi R dan minitab 16. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Membangkitkan data variabel (x) dan variabel (y).
2. Membuat diagram pencar antara variabel (x) dan variabel (y).
3. Menentukan h optimal metode Nadaraya-Watson dengan fungsi Kernel Gaussian menggunakan GCV minimum.
4. Melakukan estimasi kurva regresi nonparametrik menggunakan metode Nadaraya-Watson kernel Gaussian dengan h optimal dari GCV yang diperoleh.
5. Menguji kebaikan model dengan menghitung  $R^2$ .



## **V. KESIMPULAN**

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa estimasi kurva regresi nonparametrik menggunakan metode Nadaraya-Watson pada fungsi linear tidak dipengaruhi distribusi galat. Metode Nadaraya-Watson dengan fungsi Kernel Gaussian dapat digunakan untuk mengestimasi kurva regresi pada fungsi linear yang memiliki galat berdistribusi normal dan galat berdistribusi tidak normal.

## DAFTAR PUSTAKA

- Chatterjee, S. 2006. *Regression Analysis by Example*. Jhon Wiley and Sons, New Jersey.
- Craven, P. and 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.L. 1998. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- Halim, S. dan Bisono, I. 2006. Fungsi-Fungsi Kernel pada Metode Regresi Nonparametrik dan Aplikasinya pada Priest River Experimental Orest's Data. *Jurnal Teknik Industri*. **8**(1): 73-81.
- Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press, New York.
- 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.
- Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression*. John Wiley and Sons, New York.
- Lestari, B. dan Budiantara, I.N. 2010. Spline Estimator or Triple Response Nonparametric Regression Model. *Jurnal Ilmu Dasar*. **11**: 17-22.

- Netter, J., *et al.* 1997. *Model Linier Terapan Analisis Regresi Linier Sederhana*. Diterjemahkan oleh Bambang Sumantri. Jurusan Statistika FMIPA IPB, Bogor.
- Silverman, B. W. 1986. *Density Estimation fo Statistics and Data Analysis*. Champman & Hall, London.
- Suparti, dkk. 2018. *Regresi Nonparametrik*. Wade Group, Jawa Timur.
- Wand, M.P. and Jones, M.C. 1995. *Kernel Smoothing*. Chapman and Hall, New York.
- Yao, W. & Zhao, Z. 2013. Kernel Density-Based Linear Regression Estimate. *Communications in Statistics-Theory and Methods*. **42**: 4499-4512.