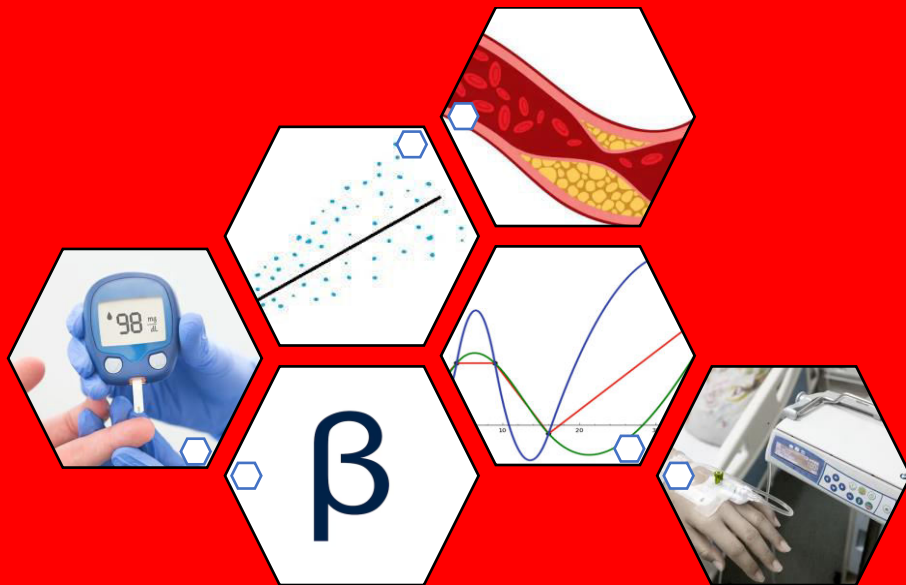


**ESTIMASI MODEL GULA DARAH PASIEN DIABETES MELALUI
REGRESI SPLINE KUADRATIK TERBOBOTI**



**PARAMITHA AR RASIS
H051171309**

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

2024



**ESTIMASI MODEL GULA DARAH PASIEN DIABETES MELALUI
REGRESI SPLINE KUADRATIK TERBOBOTI**

**PARAMITHA AR RASIS
H051171309**



**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**ESTIMASI MODEL GULA DARAH PASIEN DIABETES MELALUI
REGRESI SPLINE KUADRATIK TERBOBOTI**

PARAMITHA AR RASIS
H051171309

Skripsi



Diajukan sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

SKRIPSI

ESTIMASI MODEL GULA DARAH PASIEN DIABETES MELALUI REGRESI SPLINE KUADRATIK TERBOBOTI

PARAMITHA AR RASIS

H051171309

Skripsi,

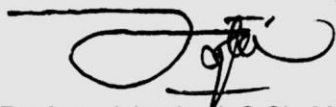
telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 2 Agustus 2024
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing tugas akhir,

Mengetahui:
Ketua Program Studi,



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul “Estimasi Model Gula Darah Pasien Diabetes Melalui Regresi Spline Kuadratik Terboboti” adalah benar karya saya dengan arahan dari pembimbing skripsi saya (Dr. Anna Islamiyati, S.Si., M.Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 2 Agustus 2024



Paramitha Ar Rasis
NIM H051171309

UCAPAN TERIMA KASIH

Alhamdulillah Robbila'lamin, Segala Puji senantiasa disenandungkan untuk menyampaikan rasa syukur penulis kepada Allah Subhanahu Wata'ala atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul "Estimasi Model Gula Darah Pasien Diabetes Melalui Regresi Spline Kuadratik Terboboti". Berkat rahmat dan izin-Nya serta dukungan dan bantuan dari berbagai pihak baik moril maupun materil, langsung maupun tidak langsung, akhirnya skripsi ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus dan penghargaan kepada:

1. Ibu Dr. Anna Islamiyati, S.Si., M.Si., selaku dosen pembimbing yang dengan sabar, tulus, dan ikhlas meluangkan begitu banyak waktu di tengah berbagai kesibukan dan prioritasnya untuk membimbing dan memberi masukan serta motivasi dalam penulisan skripsi ini.
2. Orang tua dan keluarga yang telah memberikan bantuan dukungan material dan moral; dan
3. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih untuk segala dukungan dan partisipasi yang diberikan kepada penulis, semoga bernilai ibadah di sisi Allah SWT.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga skripsi ini dapat bermanfaat bagi pihak-pihak yang berkepentingan.

Penulis,

Paramitha Ar Rasis

ABSTRAK

PARAMITHA AR RASIS. **Estimasi Model Gula Darah Pasien Diabetes Melalui Regresi Spline Kuadratik Terboboti** (dibimbing oleh Anna Islamiyati).

Latar Belakang. Diabetes Melitus (DM) Tipe II merupakan salah satu penyakit penyebab kematian terbesar keempat di dunia, salah satu tanda seseorang terkena DM yaitu naiknya kadar gula darah. Faktor yang dapat memengaruhi naiknya kadar gula darah pasien DM yaitu kolesterol. Kolesterol terdiri atas beberapa jenis, antara lain yaitu kolesterol LDL, HDL, dan kolesterol total. Pada data DM terindikasi mengalami heterokedastisitas, sehingga dalam penelitian ini akan dianalisis menggunakan regresi nonparametrik spline kuadratik dengan pembobot variansi. **Tujuan.** Penelitian ini bertujuan memperoleh estimasi fungsi spline kuadratik dengan pembobot variansi, kemudian menerapkan model yang diperoleh pada data perubahan kadar gula darah pasien DM. **Metode.** Penelitian ini menggunakan metode weighted least square (WLS). Model regresi nonparametrik spline kuadratik dengan pembobot variansi yang optimal dapat dilihat berdasarkan kriteria nilai *generalized cross validation* (GCV) minimum. **Hasil.** Berdasarkan analisis, diperoleh model spline kuadratik terboboti dengan GCV minimum sebesar 11502,03 untuk kolesterol LDL (x_1) terletak pada titik knot 59 dan 81, sedangkan untuk kolesterol total (x_2) terletak pada titik knot 99. **Kesimpulan.** Model spline kuadratik dengan pembobot variansi *error* pada data diabetes melitus tipe II berdasarkan faktor kolesterol LDL dan kolesterol total adalah $\hat{y} = 1,59 + 1,51(x_1) - 6,51(x_1)^2 - 6,32(x_1 - 59)^2_+ + 3,78(x_1 - 81)^2_+ - 3,23(x_2) - 5,00(x_2)^2 + 5,09(x_2)^2 - 6,51(x_2 - 99)^2_+$.

Kata kunci: Diabetes Melitus Tipe II, Spline Kuadratik, Heterokedastisitas, Pembobot Variansi

ABSTRACT

PARAMITHA AR RASIS. **Estimation of Blood Sugar Model of Diabetes Patients Using Weighted Quadratic Spline Regression** (supervised by Anna Islamiyati).

Background. Diabetes Mellitus (DM) Type II is one of the fourth leading causes of death in the world that is characterized by increased blood sugar level. Factors that can affect the increase in blood sugar levels of DM patients is cholesterol. Cholesterol consists of several types including LDL, HDL, and total cholesterol. DM data is indicated experiencing heteroscedasticity, so that in this study will be analyzed using nonparametric quadratic spline regression with weighting variance. **Aim.** This study aims to obtain an estimation of quadratic spline function by weighting variance, then applying the model obtained in data on changes in blood sugar levels of DM patients. **Method.** This study is using weighted least square (WLS) method. Quadratic spline nonparametric regression model with optimal variance weighting can be seen based on the criteria of the value of generalized cross validation (GCV) minimum. **Results.** Based on the analysis, obtained weighted quadratic spline models with a minimum GCV of 11502.03 for LDL cholesterol (x_1) located at knots 59 and 81 while for total cholesterol (x_2) located at knots 99. **Conclusion.** The quadratic spline model with error variance weighting on diabetes mellitus type II data based on LDL cholesterol and total cholesterol factors is $\hat{y} = 1.59 + 1.51(x_1) - 6.51(x_1)^2 - 6.32(x_1 - 59)_+^2 + 3.78(x_1 - 81)_+^2 - 3.23(x_2) - 5.00(x_2)^2 + 5.09(x_2)^2 - 6.51(x_2 - 99)_+^2$.

Keywords: Diabetes Mellitus Type II, Quadratic Spline, Heteroscedasticity, Variance Weighing

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PERNYATAAN PENGAJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Batasan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
1.5 Teori	2
1.5.1 Regresi Parametrik	2
1.5.2 Regresi Nonparametrik dengan Estimator Spline Kuadratik	3
1.5.3 Metode <i>Weighted Least Square</i>	5
1.5.4 Pemilihan Titik Knot Optimal	5
1.5.5 Heterokedastisitas.....	6
1.5.6 Diabetes Melitus Tipe II.....	6

BAB II. METODOLOGI PENELITIAN	8
3.1 Sumber Data.....	8
3.2 Variabel Penelitian.....	8
3.3 Metode Analisis	8
BAB III. HASIL DAN PEMBAHASAN.....	10
3.1 Estimasi Fungsi Spline Kuadratik Dengan Pembobot Variansi.....	10
3.2 Memodelkan Kadar Gula Darah Pasien Diabetes Tipe II dengan Estimator Spline Kuadratik Pembobot Variansi	12
3.2.1 Model Gula Darah terhadap Kolesterol LDL	13
3.2.2 Model Gula Darah terhadap Kolesterol Total.....	17
3.2.3 Model Regresi Spline Kuadratik Multivariabel	20
3.2.4 Penentuan Pembobot dan Analisis Residual Model	22
3.2.5 Model Gula Darah dengan Kolesterol LDL Terboboti	22
3.2.6 Model Gula Darah dengan Kolesterol Total Terboboti	25
3.2.7 Model Spline Kuadratik Multivariabel Terboboti dengan Knot Optimal	28
BAB IV. KESIMPULAN	31
4.1 Kesimpulan	31
4.2 Saran	31
DAFTAR PUSTAKA.....	32
LAMPIRAN	34

DAFTAR TABEL

Nomor urut	Halaman
1. Statistik deskriptif gula darah, kolesterol LDL dan kolesterol total pasien diabetes	13
2. Nilai GCV model spline kuadratik satu titik knot x_1	14
3. Estimasi model spline kuadratik satu titik knot x_1	14
4. Nilai GCV model spline kuadratik dua titik knot x_1	14
5. Estimasi model spline kuadratik dengan dua titik knot x_1	15
6. Nilai GCV spline kuadratik dengan tiga titik knot x_1	15
7. Estimasi spline kuadratik dengan tiga titik knot x_1	15
8. Nilai GCV optimal pada beberapa titik knot x_1	16
9. Nilai GCV spline kuadratik dengan satu titik knot.....	17
10. Estimasi model spline kuadratik satu titik knot x_2	18
11. Nilai GCV model spline kuadratik dua titik knot x_2	18
12. Estimasi model spline kuadratik dengan dua titik knot x_2	18
13. Nilai GCV spline kuadratik dengan tiga titik knot x_2	19
14. Estimasi spline kuadratik dengan tiga titik knot x_2	19
15. Nilai GCV optimal pada beberapa titik knot x_1	20
16. Nilai GCV spline kuadratik simultan satu titik knot	21
17. Estimasi model spline kuadratik secara simultan dengan satu titik knot	21
18. Nilai GCV model spline kuadratik terboboti satu titik knot x_1	22
19. Estimasi model spline kuadratik terboboti satu titik knot x_1	23
20. Nilai GCV model spline kuadratik terboboti dua titik knot x_1	23
21. Estimasi model spline kuadratik terboboti dua titik knot x_1	23
22. Nilai GCV spline kuadratik terboboti tiga titik knot x_1	24
23. Estimasi spline kuadratik terboboti tiga titik knot x_1	24
24. Nilai GCV terboboti optimal pada beberapa titik knot x_1	25
25. Nilai GCV spline kuadratik terboboti satu titik knot x_2	26
26. Estimasi model spline kuadratik terboboti satu titik knot x_2	26
27. Nilai GCV model spline kuadratik terboboti dua titik knot x_2	26
28. Estimasi model spline kuadratik terboboti dua titik knot x_2	27

29. Nilai GCV spline kuadratik terboboti tiga titik knot x_2	27
30. Estimasi spline kuadratik terboboti tiga titik knot x_2	27
31. Nilai GCV terboboti optimal pada beberapa titik knot x_2	28
32. Estimasi model spline kuadratik terboboti dengan knot optimal	29

DAFTAR GAMBAR

Nomor urut	Halaman
1. Plot antara gula darah pasien terhadap kolesterol LDL	13
2. Plot estimasi spline kuadratik optimal satu titik knot kolesterol LDL	16
3. Plot antara gula darah pasien diabetes terhadap kolesterol total	17
4. Plot estimasi spline kuadratik optimal dengan satu titik knot kolesterol total	20
5. Plot estimasi spline kuadratik terboboti optimal dua titik knot kolesterol LDL	25
6. Plot estimasi spline kuadratik terboboti optimal satu titik knot kolesterol total	28

DAFTAR LAMPIRAN

Nomor urut	Halaman
1. Data kadar gula darah pasien diabetes melitus dan faktor-faktor yang memengaruhi di Rumah Sakit Pendidikan Universitas Hasanuddin.	35
2. Titik knot dan Nilai GCV untuk spline kuadratik pada setiap prediktor terboboti.....	36

BAB I

PENDAHULUAN

1.1 Latar Belakang

Regresi nonparametrik merupakan salah satu pendekatan regresi yang dapat digunakan untuk melihat pola hubungan pasangan data antara variabel prediktor dengan variabel respon. Regresi nonparametrik digunakan ketika fungsi data tidak diketahui, sehingga memiliki fleksibilitas yang tinggi karena data dapat mencari sendiri bentuk estimasi fungsi regresinya tanpa dipengaruhi oleh faktor subjektivitas peneliti (Eubank, 1999). Di dalam regresi nonparametrik terdapat beberapa estimator di antaranya spline, kernel, polinomial lokal, deret fourier dan lain-lain. (Lyche dan Morken, 2004). Salah satu estimator yang sering digunakan yaitu estimator spline. Spline merupakan jenis potongan polinomial yang tersegmen sehingga memberikan fleksibilitas yang tinggi dibandingkan dengan model polinomial pada umumnya. Oleh karena itu, spline sangat efektif untuk penyesuaian bentuk fungsi data dan penanganan pola data yang berfluktuasi dengan bantuan titik-titik knot (Hardle, 1990).

Pada analisis regresi, pendekatan parametrik maupun nonparametrik memiliki asumsi uji di antaranya heterokedastisitas. Heterokedastisitas adalah ketidaksamaan nilai variansi *error* (Ghozali, 2005). Heterokedastisitas dapat dideteksi menggunakan metode grafik dan uji statistik (Gujarti, 2004). Data dikatakan mengalami heterokedastisitas jika *scatter plot* antara *error* dan variabel respon memiliki sebaran data yang tidak *random* atau membentuk pola tertentu, sedangkan uji statistik untuk heterokedastisitas terdapat beberapa, yaitu uji korelasi Rank-Spearman, uji Park, dan uji Glejser (Setawan dan Kusri, 2010).

Salah satu bentuk penanganan heterokedastisitas pada data adalah penggunaan pembobot dalam proses estimasi. Pada pendekatan regresi parametrik maupun nonparametrik, beberapa penelitian telah dilakukan dalam menangani kasus heterokedastisitas, seperti Hanifah, dkk (2015) mengatasi heterokedastisitas pada regresi linear berganda menggunakan pembobot $1/x$; Riswandi (2011) menggunakan spline terbobot pada model regresi semiparametrik dengan pembobot $1/N$; Nanik (2015) mengestimasi kurva regresi nonparametrik pada kasus heterokedastisitas dengan pembobot $1/y$.

Penelitian ini akan mengkaji kasus heteroskedastisitas dengan menggunakan pembobot variansi pada model spline kuadratik seperti yang telah dilakukan oleh Islamiyati, dkk (2018) dalam estimator penalized spline birespon yang menggunakan pembobot variansi. Estimator spline kuadratik telah dikembangkan oleh Islamiyati dan Budiantara (2007) namun masih sebatas pada data univariat yang tidak mengandung heteroskedastisitas. Selanjutnya, metode akan diterapkan pada data diabetes melitus tipe II yang diasumsikan mengalami heterokedastisitas seperti pada penelitian Dhina dan Budiantara (2011). Diabetes melitus merupakan penyakit

kronis dan telah menjadi penyebab kematian terbesar keempat di dunia. Tanda awal seseorang terkena diabetes dapat dilihat dari peningkatan kadar gula darah (Subekti, 2009). Berdasarkan uraian tersebut, penelitian ini diajukan dalam bentuk penulisan tugas akhir dengan judul **“Estimasi Model Gula Darah Pasien Diabetes Melalui Regresi Spline Kuadratik Terboboti”**.

1.2 Batasan Masalah

1. Pemilihan titik knot optimal berdasarkan nilai GCV (*Generalized Cross Validation*) yang minimum dengan 1-3 titik knot.
2. Metode estimasi yang digunakan pada penelitian ini yaitu WLS (*Weighted Least Square*).

1.3 Tujuan Penelitian

1. Mengestimasi fungsi regresi nonparametrik spline kuadratik dengan pembobot variansi pada kasus heterokedastisitas.
2. Memodelkan perubahan kadar gula darah diabetes melitus tipe II melalui estimator spline kuadratik dengan pembobot variansi.

1.4 Manfaat Penelitian

1. Penelitian ini diharapkan dapat menambah wawasan pengetahuan mengenai model regresi nonparametrik spline yang dapat mengatasi kasus heterokedastisitas.
2. Mendapatkan model kadar gula darah pasien diabetes melitus melalui spline terboboti.
3. Dapat membantu *stakeholder* pada bidang lain dalam memodelkan data yang tidak berpola parameter.

1.5 Teori

1.5.1 Regresi Parametrik

Regresi parametrik merupakan metode regresi yang digunakan ketika pola hubungan antara respon dan prediktor diketahui bentuknya, seperti linier, kuadratik atau kubik. Salah satu model regresi parametrik yang sering digunakan adalah model regresi parametrik linier berganda dengan persamaan sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \varepsilon_i ; i = 1, 2, 3, \dots, n \quad (1.1)$$

dengan

- y_i : variabel respon pada pengamatan ke- i ,
 x_{ji} : variabel prediktor ke- j pada pengamatan ke- i ,
 β_0 : intersep,

β_j : koefisien regresi ke- j ,
 ε_i : *error* pada pengamatan ke- i , dengan $\varepsilon_i \sim N(0, \sigma^2)$
(Montgomery, Peck dan Vining, 2012).

Model regresi parametrik pada persamaan (1.1) dapat dinyatakan ke dalam bentuk matriks:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.2)$$

dengan $\mathbf{y} = (y_1, y_1, \dots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, dan

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}.$$

1.5.2 Regresi Nonparametrik dengan Estimator Spline Kuadratik

Regresi nonparametrik adalah suatu metode yang digunakan untuk mengetahui pola hubungan perilaku data ketika ketersediaan informasi mengenai bentuk kurva hanya sedikit (Eubank, 1999). Kurva regresi dalam regresi nonparametrik diasumsikan halus (*smooth*) yang artinya termuat di dalam suatu ruang fungsi dan fungsi tersebut memungkinkan adanya pola data yang berubah-ubah (Budiantara, 2009). Berbeda dengan regresi parametrik, dalam regresi nonparametrik data akan mencari sendiri bentuk estimasi dari kurva regresi tanpa harus dipengaruhi oleh faktor subjektivitas peneliti (Eubank, 1999). Oleh karena itu, dalam regresi nonparametrik digunakan pendekatan yang sangat adaptif dengan persamaan regresi yang ditentukan dari data (Ryan, 1997).

Model regresi secara umum dapat ditulis sebagaimana persamaan :

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ki}) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (1.3)$$

dengan

y_i : respon pada pengamatan ke- i ,
 x_{ji} : prediktor ke- j ($j = 1, 2, \dots, k$) pada pengamatan ke- i ,
 f : fungsi/kurva regresi nonparametrik,
 n : banyaknya pengamatan,
 k : banyaknya prediktor,
 ε_i : *error* pada pengamatan ke- i , dengan $\varepsilon_i \sim N(0, \sigma^2)$

Pada model regresi nonparametrik, $f(x_{ji})$ dalam persamaan (1.3) diasumsikan tidak diketahui sehingga diestimasi dengan estimator spline. Spline adalah adalah satu jenis polinomial tersegmen yang memiliki fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik suatu fungsi atau data (Budiantara, 2006). Metode spline dalam regresi nonparametrik dapat ditemui dalam banyak bentuk, di antaranya

adalah *smoothing* spline dan *truncated* spline. Kedua metode spline masing-masing menggunakan parameter yang berbeda untuk membuat estimasi kurva regresi lebih fleksibel, yaitu parameter penghalus pada *smoothing* spline dan titik knot pada *truncated* spline. Perbedaan jenis parameter tersebut, menyebabkan optimasi untuk mendapatkan estimator pada metode spline juga berbeda (Budiantara, 2006).

Pendekatan spline umumnya menyatakan hubungan antara k prediktor dengan respon tunggal pada persamaan (1.3), dapat dilihat sebagai berikut:

$$y_i = f(x_{1i}) + f(x_{2i}) + \dots + f(x_{ki}) + \varepsilon_i; i = 1, 2, \dots, n. \quad (1.4)$$

Persamaan (1.4) dapat ditulis dalam bentuk:

$$y_i = \sum_{j=1}^k f(x_{ji}) + \varepsilon_i; i = 1, 2, \dots, n.$$

Apabila fungsi regresi f dalam persamaan (1.4) dihipotesiskan dengan fungsi *truncated* spline kuadrat yang memiliki orde $p = 2$ dan r knot, maka persamaan (1.4) dapat dinyatakan dalam bentuk sebagai berikut:

$$f(x_{ji}) = \sum_{l=0}^p \beta_{lj} x_{ji}^l + \sum_{h=1}^r \beta_{j(p+h)} (x_{ji} - K_{hj})_+^p \quad (1.5)$$

dengan

- y_i : respon pada pengamatan ke- i
- x_{ji} : prediktor ke- j pada pengamatan ke- i
- β_{0j} : intersep
- β_{lj} : parameter polinomial pada orde ke- l dan prediktor ke- j
- $\beta_{j(p+h)}$: parameter *truncated* pada prediktor ke- j dan titik knot ke- $(p+h)$
- K_{jh} : nilai titik knot pada knot ke- h dan prediktor ke- j
- r : banyaknya titik knot
- p : orde polinomial spline kuadrat
- k : banyaknya prediktor
- ε_i : *error* pada pengamatan ke- i , dengan $\varepsilon_i \sim N(0, \sigma^2)$

persamaan (1.4) untuk n data pengamatan dapat dilihat dalam bentuk matriks:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.6)$$

dengan:

$$\mathbf{y} = (y_1, y_2, \dots, y_n)'$$

$$\boldsymbol{\beta} = (\beta_{0j}, \beta_{11}, \dots, \beta_{1p}, \beta_{21}, \dots, \beta_{kp}, \beta_{1(p+1)}, \dots, \beta_{1(p+r)}, \beta_{2(p+1)}, \dots, \beta_{k(p+r)})'$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$$

$\mathbf{X} =$

$$\begin{bmatrix} 1 & x_{11} & \dots & x_{11}^p & (x_{11} - K_{11})_+^p & \dots & (x_{11} - K_{1r})_+^p & x_{21} & \dots & x_{k1}^p & (x_{21} - K_{21})_+^p & \dots & (x_{k1} - K_{kr})_+^p \\ 1 & x_{12} & \dots & x_{12}^p & (x_{12} - K_{12})_+^p & \dots & (x_{12} - K_{1r})_+^p & x_{22} & \dots & x_{k2}^p & (x_{22} - K_{21})_+^p & \dots & (x_{k2} - K_{kr})_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{1n}^p & (x_{1n} - K_{1k})_+^p & \dots & (x_{1n} - K_{1r})_+^p & x_{2n} & \dots & x_{kn}^p & (x_{2n} - K_{21})_+^p & \dots & (x_{kn} - K_{kr})_+^p \end{bmatrix}$$

Estimator $\hat{\beta}$ dapat diperoleh menggunakan metode *least square*, dengan meminimumkan jumlah kuadrat *error*:

$$\min(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \min((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \quad (1.7)$$

Apabila proses meminimumkan jumlah kuadrat *error* diselesaikan, maka akan diperoleh estimator untuk $\hat{\beta}$ adalah:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (1.8)$$

Akibatnya estimator fungsi regresinya $f(x_i)$ diberikan oleh:

$$\hat{f} = \mathbf{X}\hat{\beta} = A(k_1, k_2, \dots, k_r)\mathbf{y} \quad (1.9)$$

dengan matriks $A(k_1, k_2, \dots, k_r) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

1.5.3 Metode *Weighted Least Square*

Metode *weighted least square* merupakan prosedur yang dilakukan untuk memperoleh estimasi parameter, menentukan model yang terbaik dari suatu himpunan data dengan cara meminimumkan jumlah kuadrat *error* dalam Waode (2019). Metode *least square* yang sederhana dalam penggunaannya yaitu *Ordinary Least Square* (OLS), namun jika terjadi heterokedastisitas maka diperlukan metode *Weighted Least Square* (WLS) dengan pembobot \mathbf{W} yaitu:

$$\mathbf{W} = \begin{bmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{nn} \end{bmatrix} \quad (1.10)$$

Berdasarkan prosedur yang sama pada OLS, metode ini mendefinisikan estimasi parameter sebagai suatu nilai yang meminimumkan jumlah kuadrat *error* yang didefinisikan sebagai berikut (Searle, 1971):

$$(\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}) = ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \quad (1.11)$$

Misalkan fungsi $Q(\boldsymbol{\beta}) = ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))$, kemudian didiferensialkan terhadap $\boldsymbol{\beta}$ dan menyamakan dengan 0 maka dihasilkan parameter sebagai berikut:

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{Y}) \quad (1.12)$$

1.5.4 Pemilihan Titik *Knot Optimal*

Titik *knot* merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola data. Model regresi spline terbaik tergantung pada titik *knot optimal*. Menurut (Wang, 1998) salah satu metode yang banyak dipakai dan disukai dalam memilih titik

knot optimal adalah *Generalized Cross Validation* (GCV). Titik knot optimal diperoleh dari nilai GCV minimum. Fungsi GCV didefinisikan sebagai berikut:

$$GCV(K) = \frac{MSE(K)}{(n^{-1}tr[\mathbf{I} - \mathbf{A}(K)])^2} \quad (1.13)$$

dengan $(K) = n^{-1} \sum_{i=1}^n (y_i - \hat{y})^2$, n adalah jumlah data, \mathbf{I} adalah matriks identitas, $K = (k_1, k_2, \dots, k_r)$ adalah titik knot, dan $\mathbf{A}(K) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ merupakan matriks.

1.5.5 Heterokedastisitas

Uji Heterokedastisitas bertujuan menguji apakah dalam model regresi terjadi ketidaksamaan variansi *error* satu pengamatan ke pengamatan yang lain. Jika variansi *error* satu pengamatan ke pengamatan lain tetap, maka disebut homokedastisitas dan jika berbeda disebut heteroskedastisitas. Kebanyakan data ini menghimpun data yang mewakili berbagai ukuran (kecil, sedang, dan besar). Akibat terjadinya heteroskedastisitas maka setiap terjadi perubahan pada variabel terikat mengakibatkan *error* (residual) juga berubah sejalan atau kenaikan atau penurunannya. Dengan kata lain, konsekuensinya apabila variabel terikat bertambah maka kesalahan juga akan bertambah (Gujarati, 1998).

Uji heteroskedastisitas dilakukan dengan menggunakan Uji *Glejser*, yaitu dengan cara meregresikan antara variabel prediktor dengan nilai absolut *error*. Jika nilai signifikansi antara variabel independen dengan absolut residual lebih dari 0,05 maka tidak terjadi masalah heteroskedastisitas. Uji *Glejser* ini dinotasikan dengan $\hat{f}(x_{ji})$ pada persamaan sebagai berikut:

$$|\varepsilon_i| = \sum_{j=1}^k \hat{f}(x_{ji}) + u_i, \quad i = 1, 2, 3, \dots, n \quad (1.14)$$

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1: \text{minimal terdapat satu } \sigma_i^2 \neq \sigma^2, \quad i = 1, 2, 3, \dots, n$$

Statistik uji yang digunakan adalah sebagai berikut:

$$F_{hitung} = \frac{[\sum_{i=1}^n (|\hat{\varepsilon}_i| - |\bar{\varepsilon}|)^2] / (j - 1)}{[\sum_{i=1}^n (|\varepsilon_i| - |\hat{\varepsilon}_i|)^2] / (n - j)} \quad (1.15)$$

dengan $j = 1, 2, \dots, k$ dan $i \neq j$ dimana n adalah banyaknya variabel prediktor. Daerah penolakannya adalah tolak H_0 jika $|F_{hitung}| > F_{tabel}$, apabila H_0 ditolak artinya terdapat minimal satu $\sigma_i^2 \neq \sigma^2$ atau terjadi kasus heterokedastisitas.

1.5.6 Diabetes Melitus Tipe II

Menurut WHO, Diabetes Melitus (DM) didefinisikan sebagai suatu penyakit atau gangguan metabolisme kronis dengan multi etiologi yang ditandai dengan tingginya kadar gula darah disertai dengan gangguan metabolisme karbohidrat, lipid dan protein sebagai akibat insufisiensi fungsi insulin. Insufisiensi fungsi insulin dapat

disebabkan oleh gangguan produksi insulin oleh sel-sel beta Langerhans kelenjar pankreas, atau disebabkan oleh kurang responsifnya sel-sel tubuh terhadap insulin (Depkes, 2005). Diabetes terbagi atas beberapa jenis, salah satunya yaitu diabetes melitus tipe II. Diabetes melitus tipe II disebabkan karena sel-sel sasaran insulin gagal atau tak mampu merespon secara normal, keadaan ini biasa disebut juga "resistensi insulin". Resistensi insulin terjadi akibat obesitas, kurang aktivitas fisik, dan penuaan (Suyono, 2002).

Menurut kriteria diagnostik Perkumpulan Endokrinologi Indonesia (PERKENI), seseorang didiagnosa menderita penyakit diabetes melitus apabila memiliki kadar glukosa darah > 200 (mg/dl) dan kadar glukosa darah puasa > 126 (mg/dl), atau $HbA1C \geq 8\%$. Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, dinyatakan bahwa sosiodemografi, faktor perilaku dan gaya hidup serta keadaan klinis atau mental berpengaruh terhadap kejadian diabetes melitus (Irawan, 2010).

Berikut beberapa faktor risiko pada diabetes melitus tipe II:

a. Indeks Massa Tubuh (IMT)

Nilai dari IMT diperoleh dari pengukuran berat badan (BB) dan tinggi badan (TB). IMT dapat digunakan untuk mengetahui apakah berat badan seseorang ideal atau belum. Angka IMT yang tergolong diabetes 25,0 – 29,9 (obesitas I) dan $\geq 30,0$ (obesitas II). Hasil IMT yang masuk kategori obesitas perlu diwaspadai. Obesitas merupakan faktor risiko yang berperan penting terhadap penyakit diabetes melitus. Orang dengan obesitas mengalami kelelahan dan tidak mampu untuk memproduksi insulin yang cukup untuk mengimbangi kelebihan masukan kalori. Akibatnya kadar glukosa darah akan tinggi yang akhirnya menjadi diabetes melitus (Kaban, 2007).

b. Kadar Kolesterol

Kadar kolesterol yang tinggi berisiko terhadap penyakit DM tipe II. Kadar kolesterol tinggi menyebabkan meningkatnya asam lemak bebas sehingga terjadi lipotoksitas. Hal ini menyebabkan terjadinya kerusakan sel beta yang akhirnya mengakibatkan diabetes melitus tipe II. Kadar kolesterol total berisiko untuk diabetes jika hasilnya > 190 (mm/dL) (kolesterol tinggi), sedangkan kadar normal adalah ≤ 190 (mm/dL) (Kemenkes, 2010).

c. Tekanan Darah

Seseorang dikatakan hipertensi jika tekanan darah sistolik ≥ 140 (mmHg) atau diastolik ≥ 91 (mmHg). Hipertensi akan menyebabkan insulin resisten sehingga terjadi hiperinsulinemia, terjadi mekanisme kompensasi tubuh agar glukosa darah normal. Bila tidak diatasi maka akan terjadi gangguan Toleransi Glukosa Terganggu (TGT) yang mengakibatkan kerusakan sel beta dan terjadilah diabetes melitus tipe II (Kemenkes, 2010).

BAB II METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data penderita diabetes melitus tipe II pada tahun 2014-2018 di Rumah Sakit Pendidikan Universitas Hasanuddin.

3.2 Variabel Penelitian

Pada penelitian ini, variabel yang akan digunakan terdiri dari dua jenis, yaitu respon dan prediktor. Respon terdiri dari satu variabel, dan prediktor terdiri dari dua variabel.

a. Variabel respon (y):

Gula darah (mg/DL) adalah gula yang berada di dalam darah yang terbentuk dalam makanan dan disimpan sebagai glikogen di hati dan otot rangka.

b. Variabel prediktor (x):

x_1 adalah kolesterol LDL (mg/DL), yang bisa disebut sebagai kolesterol jahat karena kolesterol LDL menumpuk pada dinding arteri, membuat keras dan sempit.

x_2 adalah kolesterol total (mg/DL), merupakan jumlahan kolesterol yang ditemukan dalam darah, terdiri dari kolesterol LDL, kolesterol HDL, dan 20% Trigliserida.

3.3 Metode Analisis

Langkah-langkah analisis yang akan dilakukan sebagai berikut:

1. Mengestimasi fungsi regresi nonparametrik spline kuadrat dengan pembobot variansi dengan tahapan sebagai berikut:

a. Diberikan model regresi nonparametrik:

$$y_i = \sum_{j=1}^k f(x_{ji}) + \varepsilon_i; i = 1, 2, \dots, n$$

b. Menyatakan fungsi regresi $f(x_{ji})$ ke dalam fungsi spline kuadrat, $p = 2$:

$$f(x_{ji}) = \sum_{l=0}^p \beta_{lj} x_{ji}^l + \sum_{h=1}^r \beta_{j(p+h)} (x_{ji} - K_{hj})_+^p$$

dengan:

$$(x_{ji} - k_{hj})_+^p = \begin{cases} (x_{ji} - k_{hj})^p, & x_{ji} \geq k_{hj} \\ 0, & x_{ji} < k_{hj} \end{cases}$$

- c. Membuat model regresi nonparametrik spline kuadratik dalam bentuk matriks:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- d. Menentukan $\hat{\boldsymbol{\beta}}$ dengan menyelesaikan kriteria optimasi yang melibatkan pembobot \mathbf{W} untuk mengatasi heteroskedastisitas:

$$\text{Min } ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))$$

$$\text{dengan } \mathbf{W} = \begin{pmatrix} \sigma_{\varepsilon}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\varepsilon}^2 \end{pmatrix}$$

- e. Mendapatkan estimasi fungsi regresi:

$$\hat{f}(x_{ji}) = \sum_{l=0}^p \beta_{lj} x_{ji}^l + \sum_{h=1}^r \beta_{j(p+h)} (x_{ji} - K_{hj})_+^p$$

2. Memodelkan kadar gula darah pasien diabetes tipe II menggunakan estimator spline kuadratik dengan pembobot variansi, dengan 2 tahap sebagai berikut:

- Tahap I
 - a. Membuat statistik deskriptif.
 - b. Membuat *scatter plot*.
 - c. Memodelkan kadar gula darah dengan menggunakan spline kuadratik pada 1, 2, dan 3 titik knot.
 - d. Memilih titik knot optimal berdasarkan GCV minimum dari setiap titik knot pada setiap prediktor.
 - e. Memodelkan kadar gula darah (y) dengan variabel prediktor (x_1, x_2) menggunakan spline kuadratik dengan knot optimal secara simultan.
 - f. Memeriksa heteroskedastisitas data.
- Tahap II
 - g. Menghitung Pembobot \mathbf{W} .
 - h. Memodelkan kadar gula darah dengan menggunakan spline kuadratik terboboti pada 1, 2, dan 3 titik knot.
 - i. Memilih titik knot optimal berdasarkan GCV minimal dari setiap titik knot pada setiap prediktor.
 - j. Memodelkan kadar gula darah dengan variabel prediktor dengan regresi spline kuadratik terboboti dengan knot optimal secara simultan.
 - k. Menginterpretasikan hasil analisis.