

BAB II

KAJIAN TEORI

2.1 Regresi Nonparametrik

Statistik Nonparametrik adalah statistik bebas sebaran (tidak mensyaratkan bentuk sebaran populasi, baik normal ataupun tidak). Selain itu statistik nonparametrik biasanya menggunakan skala pengukuran social, seperti nominal dan ordinal yang pada umumnya tidak berdistribusi normal (Junaidi, 2015). Regresi Nonparametrik menjelaskan hubungan antara variabel prediktor terhadap variabel respon yang sebelumnya tidak diketahui fungsinya dan hanya diasumsikan mulus (Kurniasih : 2013). Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Winarti dan Sony : 2010 dalam Adzim : 2015). Regresi Nonparametrik disebut Fleksibel karena tidak memenuhi asumsi-asumsi tertentu seperti pada regresi parametrik, dan karena kurva regresi tidak diketahui data diharapkan mencari sendiri bentuk estimasinya (Esty : 2014). Filosofi utama dari model regresi nonparametrik adalah untuk mengestimasi fungsi regresi menggunakan suatu rata-rata terbobot daridata mentah dimana bobot-bobot nya merupakan suatu fungsi dari jarak dalam ruang-t (Lestari, dkk : 2019). Menurut Musholawati (2002) yang dikutip oleh Adzim (2015) Pengepasan Kurva Regresi Pembobotan tidak dilakukan pada frekuensi X tetapi pada variabel respon Y disekitar X. maka pembobot pengamatan Y_i ditentukan oleh jarak X_i terhadap X.

Regresi nonparametrik digunakan apabila bentuk pola hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuk kurva regresinya. Dalam regresi nonparametrik, kurva regresi nonparametrik hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam ruang fungsi tertentu (Indrayanti : 2016).

Secara umum model regresi nonparametrik dapat dituliskan sebagai berikut :

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

Dengan

y_i : variabel prediktor ke - i

$f(x_i)$: fungsi nonparametrik ke-I yang tidak diketahui

ε_i : error ke-i yang diasumsikan menyebar $N \sim (0; \sigma^2)$

(Yuniartika :2013)

2.2 Regresi Kernel

Regresi Kernel merupakan teknik statistika nonparametrik untuk menaksir nilai ekspetasi bersyarat dari suatu variabel random dengan nilai ekspetasi $E(Y|X)$ (Adzim:2015). Estimator Kernel merupakan suatu model nonparametrik yang umumnya digunakan , karena memiliki beberapa kelebihan yaitu :

1. Estimator Kernel mempunyai bentuk yang fleksibel dan secara matematik mudah dikerjakan.
2. Estimator kernel mempunyai rata-rata kekonvergenan yang relative cepat.

(Hardle : 1990 dalam Adzim : 2015)

Regresi kernel merupakan teknik statistika nonparametrik yang digunakan untuk mengestimasi fungsi $m(x)$ yang terdapat dalam model regresi non parametrik. Nadaraya Watson pada tahun 1994 mendefinisikan estimator regresi kernel, sehingga disebut estimator Nadaraya-Watson. Fungsi regresi $m(x)$ yang memenuhi model sebagai berikut:

$$Y_i = m(x_i) + \varepsilon_i$$

Dengan m fungsi regresi yang tidak diketahui dan ε_i variabel random independen dengan mean 0 dan varian σ^2 . Nilai $m(x)$ ekuivalen dengan nilai harapan dari variabel respon jika variabel prediktor $X = x$ telah diketahui. Secara matematik dapat ditulis dengan :

$$m(x) = E(Y|X=x)$$

diasumsikan variabel prediktor dan respon adalah variabel random, maka $m(x)$ dapat diuraikan menjadi :

$$m(x) = \int_{-\infty}^{\infty} y f(y|x) dy = \int_{-\infty}^{\infty} y f(y|x) dy = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_x(x)} dy$$

Dengan menggunakan estimator densitas kernel $f(x, y)$ dan $f_x(x)$ sebagai berikut :

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_x \left(\frac{x - X_i}{h_x} \right) K_y \left(\frac{y - Y_i}{h_y} \right) \quad (5)$$

$$\hat{f}(x, y) = \frac{1}{nh_x} \sum_{i=1}^n K_x \left(\frac{x - X_i}{h_x} \right) \quad (6)$$

Dimana

n = jumlah unit observasi

$K_x(\cdot)$ dan $K_y(\cdot)$ = fungsi kernel

h_x dan h_y = bandwidth.

X_i = data ke - i variabel prediktor

$f(x, y)$ = fungsi kernel x dan y

Sehingga $\hat{m}(x)$ diperoleh dengan mensubstitusikan persamaan 5 dan 6 :

$$\hat{m}(x) = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_x(x)} dy$$

$$\hat{m}(x) = \int_{-\infty}^{\infty} y \frac{\frac{1}{nh_x h_y} \sum_{i=1}^n K_x\left(\frac{x - X_i}{h_x}\right) K_y\left(\frac{y - Y_i}{h_y}\right)}{\frac{1}{nh_x h_y} \sum_{i=1}^n K_x\left(\frac{x - X_i}{h_x}\right)} dy \quad (7)$$

Misal $\eta = \frac{y - Y_i}{h_y}$, maka diperoleh :

$$\frac{1}{h_y} \int_{-\infty}^{\infty} y K_y\left(\frac{y - Y_i}{h_y}\right) dy = \int_{-\infty}^{\infty} (h_y \eta + Y_i) K_y(\eta) d\eta = Y_i \quad (8)$$

Dengan mensubstitusikan persamaan (8) ke dalam persamaan (7) dan untuk $h_x = h$, maka akan diperoleh estimator Nadaraya-Watson (N-W) sebagai berikut :

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \quad (9)$$

Dengan :

$$W_i(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \quad (10)$$

$\{W_i(x)\}$ = barisan bobot-bobot positif dan memiliki karakteristik

$\sum_{i=1}^n W_i(x) = 1$, maka diperoleh :

$$\hat{m}(x) = \sum_{i=1}^n W_i(x) Y_i \quad (11)$$

Sehingga estimator Nadaraya Watson merupakan rata-rata terboboti dari $\{Y_i\}$

(Suparti dkk : 2017)

Ada beberapa teknik smoothing lain yang dapat digunakan yaitu Histogram, deret orthogonal, spline, K-NN, deret fourier, dan wavelet. Estimator kernel adalah salah satu pendekatan nonparametrik dalam mengestimasi kurva yang smooth. Estimator kernel merupakan pengembangan dari estimator histogram yang diperkenalkan oleh Rosenblett (1956) dan Parzen (1962), sehingga disebut densitas kernel Rozenblett-Parzen (Hardle, 1994) (Utami : 2018). Histogram merupakan salah satu teknik yang biasa digunakan untuk mengetahui bentuk distribusi suatu data, namun tidak memberikan estimasi densitas yang mulus, sehingga dibutuhkan estimator penghalus dari fungsi densitas yang sebenarnya agar fungsi tersebut mulus dan salah satu estimator penghalus yang dapat digunakan yaitu estimator kernel yang merupakan pengembangan dari histogram (Nanda : 2016).

Selain estimator kernel, terdapat juga fungsi kernel. Suatu fungsi kernel harus merupakan fungsi kontinyu, berharga riil, simetris, dan terbatas. Secara umum fungsi kernel didefinisikan sebagai berikut didefinisikan sebagai berikut :

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right), \text{ untuk } -\infty < \infty, h > 0 \quad (6)$$

Dengan definisi $u = \frac{x - X_i}{h}$ (Ayuningtyas : 2018)

Dimana :

K : fungsi kernel

H : bandwidth atau parameter pemulus

Serta fungsi kernel di atas 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) $\int_{-\infty}^{\infty} x K(x) dx = 0$

Maka, estimator densitas kernel untuk fungsi densitas $f(x)$ adalah :

$$\hat{f}_h = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (7)$$

(Hardle : 1990)

\hat{f}_h tergantung pada fungsi kernel K dan parameter h. Bentuk bobot kernel ditentukan oleh fungsi kernel K, sedangkan ukuran bobotnya ditentukan oleh parameter pemulus h yang disebut *bandwidth*. Menurut Nyoman dan Sri (2012) (dalam Adzim, 2015), parameter pemulus (*bandwidth*) dalam fungsi densitas

kernel berfungsi untuk mengatur kehalusan kurva yang akan diestimasi. Peran bandwidth seperti lebar interval pada histogram. Ada tujuh fungsi kernel :

1. Kernel Uniform : $K(u) = \frac{1}{2}I(|u| \leq 1)$
2. Kernel Segitiga : $K(u) = (1 - |u|)I(|u| \leq 1)$
3. Kernel Epanechnikov : $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$
4. Kernel Kuartik : $K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1)$
5. Kernel Triweight : $K(u) = \frac{35}{32}(1 - u^2)^3 I(|u| \leq 1)$
6. Kernel Cosinus : $K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I(|u| \leq 1)$
7. Kernel Gaussian : $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$,
 $-\infty < x < \infty$

Dengan I adalah fungsi indikator.

$I(x)$ bernilai 1 jika $|u| \leq 1$ akan bernilai 0 jika $|u| > 1$ (Ayuningtyas : 2018)

Dari semua fungsi kernel, fungsi Kernel Gaussian yang paling sering digunakan dalam penelitian, karena fungsi kernel ini lebih smooth dibandingkan dengan fungsi kernel lain. Menurut Kurniasih (2013) fungsi kernel Gaussian lebih mudah dalam perhitungan dan penggunaannya serta lebih sering diunakan sedangkan fungsi kernel lainnya memerlukan syarat dalam pengerjaannya.

2.3 Pemilihan *Bandwidth* (h)

Bandwidth adalah parameter bebas yang menunjukkan pengaruh yang kuat pada perkiraan yang dihasilkan (Ayuningtyas : 2018). *Bandwidth* (h) adalah parameter pemulus (smoothing) yang memiliki fungsi untuk mengontrol kurva yang sedang diestimasi. Apabila nilai *bandwidth* terlalu kecil maka akan menjadi sangat kasar dan sangat fluktuatif (*under-smoothing*) Sebaliknya, jika nilai *bandwidth* (h) terlalu lebar akan menyebabkan kurva sangat halus (*over-smoothing*) karena nilai *bandwidth* terlalu lebar (Fitriani: 2015). Tujuan dari *smoothing* adalah untuk membuang variabilitas dari data yang memiliki kurva regresi tidak berpola sehingga ciri-ciri data akan terlihat lebih jelas (Esty : 2014). Oleh karena itu, diperlukan *bandwidth* (h) optimal dengan metode *Generalized Cross Validation* (GCV). *Generalized Cross Validation* merupakan salah satu metode dalam analisis regresi untuk menentukan *bandwidth* yang optimal. GCV memiliki beberapa kelebihan yaitu sederhana dalam dan efisien dalam perhitungannya, optimal secara asimtotik, invariant terhadap transformasi dan tidak memerlukan informasi terhadap σ^2 dan lebih baik digunakan pada data yang Gaussian (berdistribusi normal) (Sari : 2017). Jika Y diestimasi maka :

$$\hat{\mathbf{y}} = \mathbf{H}(\mathbf{h})\mathbf{Y} \quad (10)$$

Dimana $\mathbf{H}(\mathbf{h}) = \mathbf{W}(\mathbf{t}) + \mathbf{P}_{\tilde{\mathbf{X}}}$ dengan $\mathbf{P}_{\tilde{\mathbf{X}}} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'$

$$GCV(h) = \frac{n^{-1} \sum_{i=1}^n [y_i - \hat{y}_i]^2}{\{1 - \text{tr}(H(h)) / n\}^2}, \quad i = 1, 2, \dots, n \quad (11)$$

Dengan Keterangan :

$GCV(h)$: nilai GCV pada bandwidth h

N : banyak subjek

y_i : data actual subjek ke i

\hat{y}_i : hasil estimasi subjek ke i

$Tr(H)$: jumlah elemen dari diagonal utama matriks penghalus ukuran $n \times n$.

(Utami : 2018)

Selain itu, metode yang digunakan untuk pemilihan model terbaik adalah *Cross Validation* (CV). CV digunakan untuk pemilihan bandwidth yang *aditif*, dimana bandwidth tidak tetap tetapi bervariasi pada setiap unit observasi. Menurut Ayuningtyas (2018) dalam penelitiannya mengatakan bahwa jika bandwidth tidak tetap tetapi bervariasi tergantung pada estimasi atau sampel, hal ini akan menghasilkan kuat yang dikenal sebagai estimasi kernel bandwidth yang *aditif*, pada pemilihan bandwidth ini menggunakan metode *Cross Validation* (CV). Penggunaan metode CV dengan mempertimbangkan jumlah kuadrat residual sebagaimana bandwidth mendekati nol, mengecil $m(x)$ untuk mendapatkan Y_i yang sesuai, e_i mendekati nol dan jumlah kuadrat residual mengecil, sebagai gantinya jumlah kuadrat residual menjadi, dan merupakan kriteria yang sesuai, karena kualitas $m(x)$ dapat mempengaruhi nilai variabel independen, hal ini dapat memungkinkan terjadinya kesesuaian untuk meminimumkan variansi jumlah kuadrat residual, sehingga dinamakan

sebagai *Cross Validation* (Ayuningtyas : 2018). Dimana rumus CV adalah sebagai berikut :

$$CV(h) = n^{-1} \sum_{j=1}^n [Y_j - \hat{m}_{h,j}(X_j)]^2 \quad (12)$$

(Suparti, dkk : 2017)

Dimana :

$$\hat{m}_{h,j}(X_j) = \frac{\sum_{i \neq j} H_{ji} Y_i}{1 - H_{jj}(h)} \quad \text{untuk } i = 1, 2, \dots, n$$

Rumus CV yang lainnya adalah sebagai berikut :

$$CV = n^{-1} \sum_{j=1}^n \left[\frac{Y_j - \hat{m}_h(X_j)}{1 - H_{jj}(h)} \right]^2 \quad (13)$$

(Suparti, dkk : 2017)

Metode GCV merupakan metode CV yang terboboti, yaitu dengan memberikan bobot yang sama pada setiap pengamatan, sedangkan CV merupakan regresi terboboti dengan bobot yang berbeda-beda pada masing-masing pengamatan (Devi : 2018). Fungsi kriteria GCV memiliki pembagi yang bersifat global, sehingga baik digunakan untuk data yang memiliki variansi global yg sama, namun untuk data yang memiliki variansi populasi berbeda, lebih baik menggunakan metode CV yang memiliki bobot berbeda pada setiap pengamatan (Devi : 2018).

Nilai Bandwidth (h) yang optimal adalah nilai GCV (h) dan CV (h) yang minimum. Bandwidth optimal adalah suatu nilai bandwidth yang menghasilkan bias dan varian yang seimbang (Marlin : 2016). Untuk melihat

tingkat ukuran kesalahan estimator dengan metode *Mean Square Error* (MSE). :

$$MSE(h) = N^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

(Utami : 2018)

Apabila telah mendapatkan suatu nilai MSE yang minimum, maka nilai bias dan varian juga telah dapat diminimumkan. Cara melihat keseimbangan antara bias dan varian adalah dengan nilai MSE, yaitu dengan meminimumkan nilai MSE maka akan meminimumkan bias dan varian (Marlin : 2016). Nilai bandwidth (h) optimum menghasilkan nilai MSE terkecil. Semakin kecil nilai *Mean Square Error* (MSE) maka hasil estimasi akan semakin mendekati fungsi aslinya.

Untuk mengukur kelayakan suatu model dapat dilihat dari nilai Koefisien Determinasi atau dilambangkan dengan R^2 . R^2 merupakan besaran yang digunakan untuk mengukur kelayakan model regresi dan menunjukkan besar kontribusi X terhadap perubahan Y, nilai R^2 antara 0 sampai 1 apabila nilainya mendekati 1 maka model itu semakin baik untuk digunakan (Fitriani : 2015). Koefisien Determinasi dapat diperoleh dengan rumus :

$$R^2 = \frac{JKR}{JKT} = \frac{JKR}{JKR + JKG}$$

Dengan :

$$JKR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad JKG = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Dimana :

R^2 = koefisien determinasi

y_i = data actual subjek ke-i

\hat{y}_i = hasil estimasi subjek ke-I

\bar{y} = rata-rata dari data aktual

(Nanda : 2016)

2.4 Gizi Buruk dan Faktor-Faktor Penyebabnya

Gizi buruk merupakan suatu kondisi kekurangan gizi pada tingkatan yang sudah berat, dimana status gizinya berada jauh di bawah standar. Dimana gizi buruk akan terjadi manakala kebutuhan tubuh akan kalori, protein, atau bahkan keduanya tidak tercukupi (Ramadani, dkk: 2013). Ada tiga jenis status gizi buruk, yaitu gizi buruk karena kekurangan protein (*kwashiorkor*), kekurangan karbohidrat (*marasmus*), dan kekurangan keduanya (*marasmic-kwashiorkor*) (Ramadani, dkk: 2013). Gizi buruk berpeluang untuk menyerang siapa saja, terutama bayi dan anak-anak yang tengah berada pada masa pertumbuhan.

Faktor-faktor yang berhubungan dengan gizi buruk, Ada beberapa penyebab gizi buruk diantaranya :

1. Pemberian ASI Eksklusif

Dalam rangka menurunkan angka kesakitan dan kematian anak, *United Nation Children Fund* (UNICEF) dan *World Health Organization* (WHO) merekomendasikan sebaiknya anak hanya disusui air susu ibu (ASI) selama paling sedikit enam bulan (Kemenkes RI , 2013). Menurut WHO,

Makanan padat seharusnya diberikan sesudah anak berumur 6 bulan dan pemberian ASI dilanjutkan sampai anak berumur 2 tahun (Kemenkes, 2013). ASI eksklusif dianjurkan pada beberapa bulan pertama kehidupan karena ASI mengandung gizi yang diperlukan dan paling sesuai untuk bayi/anak dan kebersihan ASI lebih terjamin dibandingkan makanan lain (Kemenkes, 2013). Oleh karena itu pemberian ASI sangat penting untuk melindungi anak dari bahaya gizi buruk.

2. Tablet Tambah Darah

Zat besi merupakan salah satu zat gizi yang sangat dibutuhkan oleh ibu hamil. Bahkan sebelum hamil, wanita direkomendasikan untuk mencukupi kebutuhan zat besinya. Menurut penelitian yang telah dilakukan oleh Putri (2013), pemberian tablet Fe menjadi salah satu faktor yang mempengaruhi kejadian gizi buruk secara keseluruhan. Zat besi (Fe) berperan sebagai sebuah komponen yang membentuk mioglobin, yakni protein yang mendistribusikan oksigen menuju otot, membentuk enzim, dan kolagen. Selain itu, zat besi juga berperan bagi ketahanan tubuh (Kemenkes : 2018). Saat hamil, kebutuhan akan zat-zat penting yang dibutuhkan oleh tubuh tentu mengalami peningkatan. Peningkatan ini berhubungan dengan perubahan tubuh ibu, sekaligus untuk memenuhi kebutuhan gizi bayi dalam kandungan.

Menurut Kemenkes (2018), Tablet zat besi (Fe) penting bagi ibu hamil karena memiliki beberapa fungsi berikut ini:

- a. Menambah asupan nutrisi pada janin
- b. Mencegah anemia defisiensi zat besi

- c. Mencegah pendarahan saat masa persalinan
- d. Menurunkan risiko kematian pada ibu karena pendarahan pada saat persalinan

Salah satu zat gizi yang mengalami peningkatan dan sangat dibutuhkan oleh ibu hamil adalah zat besi. Ibu hamil disarankan untuk mencukupi kebutuhan zat besinya saat hamil karena jika tidak, maka dapat menyebabkan masalah pada ibu dan bayinya. Untuk perhitungan makan sebanyak 3 kali, dengan kalori sebanyak 2500 kal dapat menghasilkan 20-25 mg zat besi setiap harinya. Selama masa kehamilan lewat perhitungan 288 hari, wanita hamil bisa menghasilkan zat besi sekitar 100 mg. Dengan demikian, kebutuhan Fe (zat besi) masih kurang pada wanita hamil sehingga membutuhkan asupan tambahan berupa tablet Fe (Kemenkes : 2018).

3. Kemiskinan

Kemiskinan sekarang ini merupakan masalah yang bersifat multidimensional. Artinya karena kebutuhan manusia itu bermacam-macam, maka kemiskinan pun memiliki banyak aspek. Dilihat dari kebijakan umum, maka kemiskinan memiliki aspek primer yang berupa miskin akan aset, organisasi sosial politik, dan pengetahuan, serta keterampilan. Aspek sekunder yang berupa miskin akan jaringan sosial, sumber-sumber keuangan dan informasi. Dimensi kemiskinan tersebut terlihat dalam bentuk keuangan gizi, air, perumahan yang sehat, perawatan kesehatan yang kurang baik, dan tingkat pendidikan yang rendah (Arsyad, 2004: 237 dalam Irhamni : 2007). Kemiskinan adalah suatu kondisi ketidakmampuan secara ekonomi untuk

memenuhi standar hidup rata-rata masyarakat di suatu daerah (Anggraeny : 2016 dalam Dewi : 2018) Namun secara umum, kemiskinan adalah ketidakmampuan seseorang untuk memenuhi kebutuhan dasar standar atas setiap aspek kehidupan. Badan Pusat Statistik (2018) mendefinisikan kemiskinan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran, Jadi Penduduk Miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan.

