BAB II
TINJAUAN PUSTAKA

2.1 Diabetes Melitus

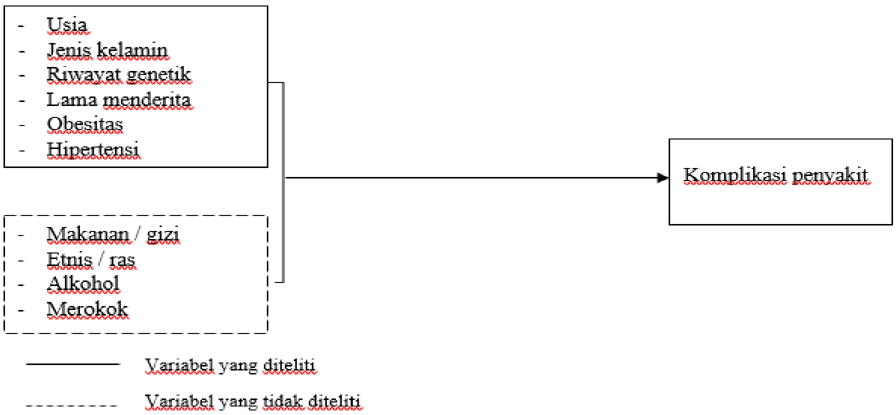
Diabetes Melitus (DM) merupakan kelompok penyakit metabolik yang ditandai dengan keadaan hiperglikemia akibat dari gangguan sekresi insulin, kerja insulin, atau keduanya (American Diabetes Association, 2014). Secara umum penyakit DM dibagi menjadi dua tipe yaitu DM tipe 1 dan DM tipe 2. DM tipe 1 merupakan penyakit autoimun dimana sistem kekebalan tubuh secara keliru menghancurkan sel beta pada pankreas yang berfungsi memproduksi insulin. Penderita DM tipe 1 biasanya adalah anak-anak dan remaja, namun terkadang juga bisa terjadi pada orang dewasa muda (Riaz, 2009).

Untuk bertahan hidup penderita DM tipe 1 memerlukan pemberian insulin setiap hari sehingga dapat menormalkan kadar glukosa dalam darah. Apabila tidak dilakukan pemberian insulin maka nyawa penderita bisa terancam dan dapat berakibat fatal. Beberapa gejala yang timbul pada penderita DM tipe 1 diantaranya rasa haus dan kencing yang ekstrim, rasa lapar yang terus menerus, penurunana berat badan, perubahan penglihatan, dan kelelahan (Okur, Karantas, dan Siafaka, 2017). DM tipe 2 merupakan suatu gangguan metabolisme yang melibatkan resistensi insulin. Pada penderita DM tipe 2, pankreas masih mampu memproduksi insulin, tetapi tubuh mengalami kesulitan menggunakan hormon pengontrol glukosa ini. Akibatnya pankreas tidak dapat memproduksi insulin yang cukup untuk menanggapi kebutuhan tubuh akan hal itu. Sejauh ini, DM tipe 2 merupakan tipe DM yang paling umum, mencakup 85 hingga 95% kasus DM. Berbeda dengan DM tipe 1 penderita DM tipe 2 tidak membutuhkan pemberian insulin, tetapi hanya memerlukan obat untuk memperbaiki fungsi insulin, menurunkan gula, dan lain-lain Penyakit DM tipe 2 membutuhkan waktu bertahun-tahun untuk berkembang. Biasanya didahului oleh prediabetes, dimana kadar glukosa darah di atas normal tetapi belum cukup tinggi untuk terdiagnosis diabetes (Riaz, 2009).

Pengontrolan kadar gula darah pada penderita DM sangat penting untuk dilakukan. Semakin buruk kontrol gula darah maka semakin mudah seseorang terkena komplikasi penyakit. Komplikasi akibat penyakit DM dapat dibagi dalam dua kelompok yaitu komplikasi akut dan komplikasi kronis. Komplikasi akut biasanya terjadi secara mendadak. Beberapa komplikasi penyakit akut akibat diabetes yaitu hipoglikemia, ketoasidosis diabetic, dan koma hiper osmolar non-ketotik. Berbeda dengan komplikasi akut, komplikasi kronis biasanya timbul secara perlahan, kadang tidak diketahui, tetapi akhirnya berangsur menjadi makin berat dan membahayakan. Beberapa komplikasi penyakit kronis yaitu kerusakan saraf (neuropati), kerusakan ginjal (*nephropathy*), kerusakan mata, penyakit jantung, stroke, penyakit pembuluh darah perifer, gangguan pada hati, penyakit paru-paru, gangguan saluran makan, infeksi, dan penyakit kulit (Tandra, 2018).

2.2 Faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2

Komplikasi penyakit DM tipe 2 dapat disebabkan oleh beberapa faktor. Berikut ini merupakan kerangka konsep mengenai faktor yang mempengaruhi penyakit DM tipe 2 berdasarkan penelitian yang telah dilakukan oleh Hasanah (2018).



Gambar 2.1 Kerangka Konsep Faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2

Berikut ini merupakan teori mengenai faktor-faktor yang mempengaruhi komplikasi penyakit DM tipe 2.

1. Usia

Salah satu faktor yang mempengaruhi komplikasi pada penyakit diabetes mellitus adalah usia. Risiko untuk menderita intoleransi glukosa meningkat seiring dengan meningkatnya usia. Artinya semakin bertambah usia seseorang maka semakin besar risiko terjadinya komplikasi pada penyakit diabetes mellitus (PERKENI, 2015).

2. Jenis Kelamin

Jenis kelamin merupakan salah satu faktor risiko diabetes yang tidak dapat diubah. Berdasarkan penelitian yang dilakukan oleh (Wahyuni, 2010) diperoleh hasil bahwa jenis kelamin perempuan memiliki kecendrungan 1,39 kali untuk terkena penyakit DM dibandingkan dengan jenis kelamin laki-laki.

3. Riwayat Genetik

Riwayat genetik juga merupakan salah satu faktor risiko diabetes yang tidak dapat diubah. Jika salah satu dari orang tua terkena diabetes maka resiko untuk menderita diabetes adalah sebesar 15%. Jika kedua orangtua terkena diabetes maka resiko untuk menderita diabetes adalah sebesar 75% (Diabetes UK, 2010).

4. Lama Menderita

Menurut (Bansal, et al., 2014) risiko pasien yang menderita DM selam 6-10 tahun untuk terkena penyakit komplikasi Neuropati Diabetik Perifer (NDP) adalah sebesar 1,33 kali, sedangkan pasien yang menderita DM lebih dari 15 tahun mempunyai risiko terkena komplikasi NDP sebesar 8,03 kali. Hal ini menunjukkan bahwa semakin lama seseorang menderita DM maka risiko terjadinya komplikasi pada penyakit DM akan semakin besar.

5. Obesitas

Obesitas merupakan suatu kelainan yang ditandai oleh penimbunan jaringan lemak dalam tubuh secara berlebihan. Obesitas akan menyebabkan resistensi insulin sehingga insulin tidak dapat bekerja dengan baik dan kadar gula darah bisa naik. Berdasarkan PERKENI (2015) salah satu faktor risiko terjadinya penyakit DM adalah obesitas (IMT $\geq$ 23 kg/m^2)

6. Hipertensi

Hipertensi biasanya terjadi bila tekanan darah mencapai lebih dari 140 mmHg (sistolik) dan lebih dari 90 mmHg (diastolik) pada dua kali pengukuran selang waktu lima menit dalam keadaan cukup istirahat (Kemenkes, Infodatin: Hipertensi, 2014). Hipertensi dapat menyebabkan penebalan pembuluh darah arteri sehingga diameter pembuluh darah menjadi menyempit. Penyempitan diameter tersebut akan menyebabkan proses pengangkutan glukosa dari dalam darah menjadi terganggu sehingga akan berisiko terhadap penyakit DM.

2.3 Regresi Logistik Biner

2.3.1 Model Regresi Logistik Biner

Analisis regresi logistik merupakan salah satu metode regresi yang dapat digunakan untuk menggambarkan hubungan variabel terikat (Y) yang bersifat kategorik dengan satu atau lebih variabel bebas (X) yang bersifat kontinu, kategori atau kombinasi keduanya (Agresti, 2002:165). Analisis ini bertujuan untuk melihat kemungkinan atau probabilitas dari suatu kejadian dengan data fungsi logit dari kurva logistik. Regresi ini juga bertujuan untuk menanggulangi kelemahan dari regresi linear sederhana karena regresi logistik tidak mengasumsikan adanya hubungan yang linier antara variabel dependen dan independennya.

Berdasarkan jumlah kemungkinan yang dihasilkan oleh variabel dependennya, regresi logistik dibedakan atas dua jenis yakni regresi logistik multinomial dan regresi logistik biner. Regresi logistik multinomial digunakan apabila variabel dependen menghasilkan lebih dari dua kemungkinan, misalnya kemungkinan yang dihasilkan dalam suatu pertandingan “menang”, “kalah” dan “seri”. Sedangkan analisis regresi logistik biner merupakan analisis regresi yang dapat digunakan apabila variabel dependennya hanya memiliki dua kemungkinan nilai, misalnya sukses dan gagal (Montgomery, 2006:428). Jika $Y = 1$ menyatakan bahwa suksesnya suatu kejadian, maka $Y = 0$ menyatakan bahwa kejadian tersebut gagal sehingga peluang tiap kemungkinan dapat dinyatakan sebagai berikut:

$$P(Y_i = 0) = 1 - \pi_i$$

$$P(Y_i = 1) = \pi_i$$

maka model regresi peubah respon biner yang melibatkan transformasi logistik dinamakan dengan model regresi logistik, dirumuskan sebagai berikut:

$$\mathrm{P}(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \qquad (2.1)$$

Untuk mempermudah penaksiran parameter dan untuk mendapatkan fungsi yang linier maka akan dilakukan transformasi pada persamaan (2.1), dengan menggunakan transformasi logit. Uraian transformasi tersebut adalah sebagai berikut:

$$\mathrm{P}(x) = \frac{exp(\beta_0+\beta_1 x)}{1+exp(\beta_0+\beta_1 x)} = \{P(x)\}\{1 + exp(\beta_0 + \beta_1 x)\} = exp(\beta_0 + \beta_1 x)$$

$$\{P(x)\} + \{P(x) exp(\beta_0 + \beta_1 x)\} = exp(\beta_0 + \beta_1 x)$$

$$P(x) = exp(\beta_0 + \beta_1 x) - P(x)\, exp(\beta_0 + \beta_1 x)$$

$$P(x) = \{1 - P(x)\}\, exp(\beta_0 + \beta_1 x)$$

$$\frac{P(x)}{1-P(x)} = exp(\beta_0 + \beta_1 x)$$

$$ln\left[\frac{P(x)}{1-P(x)}\right] = ln\,\{exp(\beta_0 + \beta_1 x)\}$$

$$ln\left[\frac{P(x)}{1 - P(x)}\right] = \beta_0 + \beta_1 x$$

$$g(x) = \beta_0 + \beta_1 x$$

Hasil transformasi di atas maka diperoleh $g(x) = \beta_0 + \beta_1 x$ merupakan bentuk logit. Sedangkan untuk model regresi logistik biner dengan k variabel prediktor adalah sebagai berikut:

$$\pi(x) = \frac{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)} \tag{2.2}$$

dimana :

X_i	: variabel independen,	i = 1, 2, ... , k
Y	: variabel dependen,	
β_0	: penduga parameter awal	
$\beta_1, \beta_2, \ldots, \beta_k$	: penduga parameter 1,2,.....,k	
k	: banyaknya variabel independen	

Pelanggaran asumsi normalitas dan ragam galat yang tidak homogen dapat dilihat dengan menggunakan model peluang linier, yang didefinisikan sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Dengan nilai Y = 0 atau 1

Pada regresi linear, ε diasumsikan menyebar menurut sebaran normal dengan nilai tengah nol, yang berarti E(ε) = 0, sehingga dari persamaan di atas diperoleh

$$E(Y) = E(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon)$$

$$= E(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_K) + E(\varepsilon)$$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{2.3}$$

Jika P (Y = 1 $|x$) = P(x) dan P(Y = 0 $|x$) = 1 − P(x), maka berdasarkan defenisi nilai harapan diperoleh

$$E(Y) = 0\,(1 - P(x)) + 1(P(x))$$

$$= \pi\,(x) \qquad (2.4)$$

Dari persamaan 2.3 dan 2.4 diperoleh

$$\pi\,(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \qquad (2.5)$$

Terpenuhi atau tidaknya asumsi kenormalan sisaan ε dapat diketahui berdasarkan persamaan berikut:

$$\varepsilon = Y - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

$$= Y - E(Y)$$

$$= Y - \pi\,(x) \qquad (2.6)$$

Jika $Y = 1$ maka $\varepsilon = 1 - P(x)$ dan jika $Y = 0$ maka $\varepsilon = 0 - P(x) = -P(x)$. Dengan demikian ε tidak dapat menyebar secara normal karena hanya memiliki 2 kemungkinan nilai.

Asumsi kehomogenan ragam dapat diketahui dari persamaan berikut:

$$var(\varepsilon) = E(\varepsilon^2) - \left(E(\varepsilon)\right)^2 = E(\varepsilon^2) - 0 = E(\varepsilon^2)$$

$$= \left(1 - P(x)\right)^2 P(Y = 1|x) + \left(-P(x)\right)^2 P(Y = 0|x)$$

$$= \left(1 - P(x)\right)^2 P(x) + P(x)^2\left(1 - P(x)\right)$$

$$= \left(1 - P(x)\right)P(x)\left(\left(1 - P(x)\right) + P(x)\right)$$

$$= \left(1 - P(x)\right)P(x) \qquad (2.7)$$

Dari persamaan (2.7) di atas, dapat disimpulkan bahwa ragam galat ε tidak homogen karena masing-masing pengamatan mempunyai ragam galat yang berbeda. Nilai ragam galat tergantung pada P(x), dengan nilai P(x) yang berbeda akan diperoleh ragam galat yang berbeda pula. Sehingga asumsi keragaman galat tidak terpenuhi.

Selanjutnya melakukan tranformasi logit terhadap P(x). Jika respon Y yang mempunyai nilai 1 dengan peluang $\mathrm{P}(Y = 1|X = x) = \mathrm{P}(x)$ dan g(x) adalah model regresi linear

$$\text{g(x)} = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

maka model regresi peubah respon biner yang melibatkan transformasi logistik dinamakan dengan model regresi logistik, dirumuskan sebagai berikut:

$$\mathrm{P}(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \tag{2.8}$$

Tranformasi ini bertujuan untuk menyatakan bahwa regresi memenuhi sifat linear pada parameternya. Sehingga diperoleh persamaan dari tranformasi sebagai berikut:

$$\mathrm{P}(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

$$= \frac{1}{1 + e^{-g(x)}}$$

$$\mathrm{P}(x)\left(1 + e^{-g(x)}\right) = 1$$

$$e^{-g(x)} = \frac{1 - P(x)}{\mathrm{P}(x)}$$

$$e^{g(x)} = \frac{\mathrm{P}(x)}{1 - \mathrm{P}(x)}$$

$$\ln e^{g(x)} = ln\left(\frac{\mathrm{P}(x)}{1 - \mathrm{P}(x)}\right)$$

$$g(x) = logit\ (\mathrm{P}(x))$$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \qquad (2.9)$$

g(x) diatas merupakan penduga logit yang berperan sebagai fungsi linear dari peubah penjelas. Hosmer (1989:31) menyebutkan karena fungsi penghubung yang digunakan adalah fungsi penghubung logit maka sebaran peluang yang digunakan disebut sebaran logistik.

2.3.2. Estimasi Parameter

Metode *Maximum Likelihood Estimator* (MLE) digunakan untuk mengestimasi parameter-parameter dalam regresi logistik yang pada dasarnya metode ini memberikan nilai estimasi β dengan memaksimumkan fungsi likelihoodnya. Fungsi likelihood menjelaskan peluang data pengamatan sebagai fungsi parameter yang belum diketahui, sehingga sebelum menduga parameter logistik kita ketahui dulu fungsi likelihood.

Secara matematis fungsi likelihood dapat ditulis dengan:

$$L(\beta) = \prod_{i=1}^{n} f(\beta, y) \qquad (2.10)$$

Menurut Hosmer (1989:8), jika Y dikotomus memiliki dua kemungkinan 0 atau 1, maka ekspresi P(x) menghasilkan Y dengan syarat X. Jika Y=1 dinyatakan dengan P(Y=1|x) dan Y=0 dinyatakan dengan P(Y=0|x) = 1 - P(X=x). Sehingga untuk pasangan (x_i, y_i), dimana menurut fungsi likelihood $y_i = 1$ kontribusinya P(x_i) dan $y_i = 0$ kontribusinya 1-P(x_i). Dimana P(x_i) menyatakan nilai P(X=x) yang dihitung saat x = x_i. Sehingga fungsi likelihood untuk (x_i, y_i) dinyatakan dengan rumus:

$$L(\beta) = \prod_{i=1}^{n} f(\beta, y) = \prod_{i=1}^{n} P(xi)^{y_i} (1 - P(xi))^{1-y_i} \qquad (2.11)$$

Fungsi likelihood sebagai fungsi log disebut fungsi log likelihood. Fungsi likelihood untuk regresi logistik dinyatakan sebagai berikut:

L (**β**) = *ln* [L(β)]

Substitusi ke persamaan (2.5), diperoleh:

L (**β**) = *ln* {$\prod_{i=1}^{n} P(xi)^{y_i}$ (1- $P(x_i))^{1-y_i}$}

Kemudian masukan jumlah i untuk masing-masing variabel (x_i, y_i),

dimana i = 1,,,n.

L (**β**) = *ln*{[$P(x_1)^{y_1}$(1-$P(x_1))^{1-y_1}$]. [*ln* {$P(x_2)^{y_2}$(1- $P(x_2))^{1-y_2}$] ...

[*ln*{$P(x_n)^{y_n}$(1-$P(x_n)^{1-y_n}$]}

Sehingga diperoleh fungsi likelihood sebagai berikut:

L(**β**) = $\sum_{i=1}^{n}$\{ $y_i lnP(x_i)$ + (1−y_i) *ln* [1-$P(x_i)$]}

Jadi, fungsi likelihood pada regresi logistik adalah:

L(**β**) = $\sum_{i=1}^{n}$\{ $y_i lnP(x_i)$ + (1−y_i) *ln* [1-$P(x_i)$]} (2.12)

Maksimum likelihood merupakan nilai penduga parameter dengan memaksimumkan fungsi log likelihood. Dengan mendiferensialkan bentuk log likelihood terhadap $\beta_1, \beta_2, \ldots, \beta_k$ dan menyamakan dengan nol, sehingga diperoleh:

$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_i} = 0$; i= 0,1,…,k

$$\frac{\partial}{\partial \beta_i} \mathrm{L}(\beta) = \sum_{i=1}^{n} [y_i \, ln \, P \, (x_i) + (1 - y_i) ln(1 - P(x_i))] = 0$$

Sehingga persamaan penduga parameter regresi logistik sebagai berikut:

$\frac{\partial L(\beta)}{\partial \beta_i} = \sum_{i=1}^{n}[y_i$-P($x_i$)] =

(13)

$\frac{\partial L(\beta)}{\partial \beta_i} = \sum_{i=1}^{n} x_i \, [y_i$ - $P(x_i)$] = 0 ; i= 1,2,…,k (2.13)

Menurut Montgomery (2006:441) metode maksimum likelihood adalah suatu metode untuk mengestimasi parameter pada suatu persamaan dengan memaksimumkan nilai L(β) atau disebut dengan *conditional log-likelihood function* yang berasal dari probabilitas persamaan regresi logistik yang akan diestimasinya. Untuk mencari *conditional log-likelihood function* yang maksimum pada maksimum likelihood dapat menggunakan metode Newton Raphson.

Metode Newton Raphson merupakan metode untuk menemukan akar dari persamaan dengan asumsi f(x) = 0. Bentuk persamaan dari metode Newton Raphson untuk menentukan maksimum likelihood yang berasal dari turunan pertama dan kedua dari *conditional log-likelihood*. Turunan kedua dari *conditional log-likelihood* sebagai berikut:

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial {\beta_j}^2} = -\sum_{i=1}^{n} {x_{ij}}^2 \, P(x_i)[y_i - P(x_i)]$$

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_u} = -\sum_{i=1}^{n} x_{ij} x_{iu} \, P(x_i)[y_i - P(x_i)]$$

Untuk memperoleh nilai estimasi parameter yang paling optimal adalah sebagai berikut:

$$\beta_{t+1} = \beta_t + ((X'VX)^{-1} X'(y - P(x_i)) \tag{2.14}$$

Iterasi akan berhenti apabila nilai $\beta_t = \beta_{t+1}$, jika nilai $\beta_t \neq \beta_{t+1}$ maka iterasi dilanjutkan dan kembali. Dimana t merupakan tahapan iterasi, X merupakan matriks berukuran (nxk) berisi data masing-masing individu pengamatan dan V matriks diagonal berukuran (nxn) yang nilai umumnya diagonal ke-i nya adalah $P(x_i)(1 - P(x_i)$. Sedangkan nilai varian (β_j) adalah unsur diagonal ke-j dari matriks invers $I^{-1}(\beta)$ =(X' VX).

2.3.3. Uji Signifikansi Parameter

Menurut Hosmer dan Lemeshow (2000), uji signifikansi parameter yang digunakan adalah uji rasio Likelihood dan uji Wald.

a. Uji Rasio Likelihood

Uji Rasio Likelihood adalah uji signifikansi parameter secara keseluruhan atau bersama sama.

Hipotesis :

H_0: $\beta_1 = \beta_2 = \cdots = \beta_j = 0$

H_1: paling sedikit ada satu $\beta_i \neq 0$, untuk i = 1,2,3,...,k

Dengan statistika uji

$$G = 2ln\left[\frac{likelihood\ tanpa\ variabel\ penjelas}{Likelihood\ dengan\ variabel\ penjelas}\right]$$

Kriteria uji : Tolak H_0 G > $\chi^2_{a,k}$

b. Uji Wald

Uji wald digunakan untuk mengetahui apakah masing-masing variabel prediktornya memiliki pengaruh terhadap model atau tidak.

Hipotesis:

H_0: $\beta_j = 0$, untuk $j = 1,2,\ldots,k$ (peubah X_j tidak berpengaruh nyata)

H_1: $\beta_j \neq 0$ (peubah X_j berpengaruh nyata)

Dengan statistik uji :

$$W_j = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \qquad ; j = 0,1,2,\ldots,k$$

Tolak H_0 jika $W > Z\alpha/_2$ atau nilai signifikansi kurang dari α.

2.3.4 Uji Kesesuaian Model

Pengujian kesesuaian model dilakukan menggunakan *Hosmer-Lemeshow goodness-of-fit test* dengan hipotesis sebagai berikut :

H_0: model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

H_1: model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

Statistik uji :

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Dimana :

o_k : observasi pada grup ke-k ($\sum_{j=1}^{c_k} y_j$ dengan c_k: respon (0,1))

$\bar{\pi}_k$: rata – rata taksiran peluang ($\sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$)

g : jumlah grup (kombinasi kategori dalam model serentak)

n'_k : banyak observasi pada grup ke-k

Daerah kritis :

Tolak H_0 $X^2{}_{hitung} > X^2{}_{(db,\alpha)}$

Penentuan klasifikasi menggunakan territorial map yang merupakan output dari software yang digunakan. APER (Apparent Error Rate) merupakan bagian pengamatan yang mengalami kesalahan klasifikasi menurut fungsi klasifikasi. Tingkat kesalahan dapat dihitung dari confusion matrix yang menunjukkan keanggotaan kelompok aktual dan prediksi. Contoh untuk n1 dari grup 1 dan n2 dari grup 2, bentuk confusion matrix sebagai berikut (Johnson dan Winchern, 2007).

Tabel 2.1 Ketepatan Klasifikasi

	Grup 1	Grup 2
Grup 1 (n_1)	n_{1c}	$n_{1M} \square n_1 \square n_{1c}$
Grup 2 (n_2)	$n_{2M} \square n_2 \square n_{2c}$	n_{2c}

Dimana :

$n_{1c} = n_{2c}$ = jumlah anggota grup 1/ grup 2 yang diklasifikasikan benar sebagai grup 1/ grup 2

$n_{1m} = n_{2m}$= jumlah anggota grup 1/ grup 2 yang diklasifikasikan salah sebagai grup 1/ grup 2

APER = $\frac{n_{1m}+n_{2m}}{n_1+n_2}$

2.4 Multivariate Adaptive Regression Spline (MARS)

MARS merupakan salah satu pendekatan regresi nonparametrik multivariat yang berguna untuk mengatasi permasalahan data yang berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebanyak $3 \leq n \leq 20$ (Friedman, 1991). Selain itu model MARS mampu menghasilkan prediksi variabel respon yang akurat. Metode MARS menjadi popular karena tidak menentukan tipe khusus seperti hubungan (linear, kuadratik, dan kubik) diantara variabel prediktor dan respon pada proses pembentukan model MARS tidak memerlukan asumsi (Otok, Guritno, Subanar, dan Haryatmi, 2006).

Beberapa istilah yang perlu diperhatikan dalam pemodelan MARS adalah sebagai berikut.

1. Knot

Knot merupakan suatu nilai/titik tempat perubahan pola apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data yang ada dari variabel prediktor. Knot merupakan akhir dari sebuah garis regresi dan juga awal dari garis regresi yang lain (Nash dan Bradford, 2001).

2. Basis Fungsi

Basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Basis fungsi bisa memiliki lebih dari satu variabel yang merupakan fungsi dari tiap garis regresi yang dihasilkan. Maksimum basis fungsi yang diperbolehkan adalah 2 sampai 4 kali jumlah variabel prediktornya.

3. Interaksi

Interaksi merupakan hubungan korelasi antar variabel dengan maksimum interaksi (MI) adalah 1, 2, dan 3. Jika MI lebih dari tiga maka akan menghasilkan model yang lebih kompleks.

4. Minimum Observasi

Minimum Observasi (MO) merupakan jumlah pengamatan paling minimal antar knot sebesar 0, 1, 2, dan 3.

Secara umum estimator model MARS dapat ditulis pada persamaan berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{m=1}^{M} \hat{\alpha}_m \prod_{k=1}^{K_m} \left[S_{km} \left(X_{v(km)} - t_{km} \right) \right] \qquad (2.15)$$

dengan :

$\hat{\alpha}_0$ = koefisien konstanta basis fungsi B_0

α_m= koefisien dari basis fungsi ke-*m*

M = maksimum basis fungsi

m = banyaknya basis fungsi

k_m= banyaknya interaksi pada basis fungsi *m*

k = banyaknya interaksi

S_{km}= nilainya 1 atau -1 jika data berada di sebelah kanan atau kiri titik knot

$X_{v(km)}$= variabel prediktor

t_{km}= nilai knot dari variabel prediktor $x_{v(k,m)}$

Berdasarkan persamaan (2.15) model MARS dapat dituliskan sebagai berikut.

$$y_i = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_{im}(x) + \varepsilon_i \qquad (2.16)$$

dengan. $B_{im}(x) = \prod_{k=1}^{K_m} [S_{km}(X_{v(k,m)i} - t_{km})]$. Sehingga jika ditulis dalam bentuk matriks dapat menjadi

$$y = B\,\alpha + \varepsilon \qquad 2.17)$$

dengan,

$$y = (y_1 y_2 y_3, \ldots, y_n)^T, \alpha = (\alpha_1 \alpha_2 \alpha_3, \ldots, \alpha_M)^T, \varepsilon = (\varepsilon_1 \varepsilon_2 \varepsilon_3, \ldots, \varepsilon_n)^T$$

$$\boldsymbol{B} = \begin{pmatrix} 1 & \prod_{k=1}^{K_1}\left[S_{k1}\left(x_{v(k,1)1} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_m}\left[S_{kM}\left(x_{v(k,M)1} - t_{kM}\right)\right] \\ 1 & \prod_{k=1}^{K_1}\left[S_{k1}\left(x_{v(k,1)2} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_m}\left[S_{kM}\left(x_{v(k,M)2} - t_{kM}\right)\right] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1}\left[S_{k1}\left(x_{v(k,1)n} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_m}\left[S_{kM}\left(x_{v(k,M)n} - t_{kM}\right)\right] \end{pmatrix}$$

Persamaan (2.15) dapat dijabarkan sebagai berikut.

$$\hat{f}_{(X)} = \hat{\alpha}_0 + \sum_{m=1}^{M} \alpha_m \left[S_{1m}\left(X_{v(1m)} - t_{1m}\right)\right] + \sum_{m=1}^{M} \alpha_m \left[S_{1m}(X_{1m} - t_{1m})\right] + \left[S_{2m}\left(X_{v(2m)} - t_{2m}\right)\right] + \sum_{m=1}^{M} \alpha_m \left[S_{1m}\left(X_{v(1m)} - t_{1m}\right)\right] + \left[S_{2m}\left(X_{Vv2m)} - t_{2m}\right)\right] + \left[S_{3m}\left(X_{v(3m)} - t_{3m}\right)\right] + \cdots \quad (2.18)$$

Secara umum persamaan (2.18) dapat dituliskan sebagai berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{i=1}^{v} f_i(x_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{v} f_{ij}\left(x_i, x_j\right) + \sum_{\substack{i,j=1 \\ i \neq j}}^{v} f_{ijk}\left(x_i, x_j, x_k\right) + \ldots \quad (2.19)$$

Persamaan (2.19) menunjukkan bahwa penjumlahan suku pertama meliputi semua basis fungsi untuk satu variabel. Penjumlahan suku kedua meliputi semua basis fungsi untuk interaksi antara dua variabel. Penjumlahan suku ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel, dan seterusnya.

Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai GCV yang minimum (Nisa' dan Budiantara, 2012). Pemilihan model pada MARS dapat menggunakan metode *stepwise* (*forward* dan *backward)*. Pemilihan model dengan menggunakan *forward stepwise* dilakukan untuk mendapatkan jumlah basis fungsi maksimum, sedangkan pada *backward stepwise* dilakukan pemilihan basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross*

Validation (GCV). Model terbaik dipilih berdasarkan nilai GCV yang paling minimum. Fungsi GCV minimum didefenisikan sebagai berikut.

$$GCV\ (M) = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{f}_m(x_i)^2]}{\left[1 - \frac{C(M)}{N}\right]^2} \tag{2.20}$$

dengan,

N : banyaknya pengamatan

y_i : variabel respon

x_i : variabel prediktor

$\hat{f}_m(x_i)$: nilai taksiran variabel respon pada pengamatan ke-i

$c(M)$: jumlah parameter dalam model = $Trace\ [B(B^T B)^{-1} B^T] + 1$

B : matriks basis fungsi

2.4.1 Klasifikasi MARS Respon Biner

Klasifikasi pada model MARS didasarkan pada analisis regresi. Jika variabel respon terdiri dari dua nilai, maka dikatakan sebagai regresi dengan respon biner (Cox dan Snell, 1989), sehingga dapat digunakan model probabilitas dengan persamaan sebagai berikut.

$$P(Z = 1\ |X = x) = \pi(x) = \frac{e^{f(x)}}{1 + e^{f(x)}}$$

dan

$$P(Z = 0\ |X = x) = 1 - \pi(x) = \frac{1}{1 + e^{f(x)}}$$

Dengan $f\ (x) = z = logit\ \pi\ (x)$. Model MARS untuk klasifikasi dapat dinyatakan sebagai berikut.

$$z = logit\ \pi\ (x) = \ln\left(\frac{\pi(x)}{1-\ \pi(x)}\right)$$

$$= \ \alpha_0 + \sum_{m=1}^{M} \alpha_0 \prod_{k=1}^{Km} \left[S_{km}\left(x_{v(km)} - t_{km}\right)\right] \quad (2.21)$$

2.4.2 Ketepatan Klasifikasi

Ketepatan klasifikasi diperlukan untuk mengetahui pengelompokkan data yang digolongkan dengan tepat pada kelompoknya. *Apparent Error Rate* (APER) didefenisikan sebagai proporsi sampel yang tidak tepat diklasifikasikan (Johnson dan Wichern, 2007). Untuk mengetaui proporsi sampel yang tepat diklasifikasikan dapat dihitung dari nilai TAR *(Total Accuracy Rate)*. Berikut ini merupakan tabel pengklasifikasian untuk respon biner.

Tabel 2.1 Klasifikasi Respon Biner

Observasi	Taksiran Observasi	
	y_0	y_1
y_0	n_{00}	n_{01}
y_1	n_{10}	n_{11}

Nilai APER dan TAR didapatkan dengan perhitungan sebagai berikut.

$$APER(\%) = \frac{n_{10}+n_{01}}{n} X\ 100\% \quad (2.22)$$

dan

$$TAR\ (\%) = 1 - APER = 1 - \left(\frac{n_{10}+n_{01}}{n} X\ 100\%\right) \quad (2.23)$$

dengan,

n = jumlah observasi

n_{00} = jumlah observasi dari y_0 yang tepat diklasifikasikan sebagai y_0

n_{11} = jumlah observasi dari y_1 yang tepat diklasifikasikan sebagai y_1

n_{01} = jumlah observasi dari y_0 yang salah diklasifikasikan sebagai y_1

n_{10} = jumlah observasi dari y_1 yang salah diklasifikasikan sebagai y_0