

BAB II

TINJAUAN PUSTAKA

2.1 Tenaga Kerja

Tenaga Kerja adalah penduduk yang berada dalam usia kerja. Menurut Undang-Undang Nomor 13 Tahun 2003 tentang ketenagakerjaan, yang disebut sebagai tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang dan atau jasa baik untuk memenuhi kebutuhan sendiri maupun masyarakat. Jadi, tenaga kerja merupakan semua orang yang bersedia untuk sanggup bekerja tanpa batasan umur tertentu. Tenaga kerja meliputi mereka yang bekerja untuk diri sendiri ataupun anggota keluarga yang tidak menerima bayaran berupa upah atau mereka yang sesungguhnya bersedia dan mampu untuk bekerja, dalam arti mereka menganggur dengan terpaksa karena tidak ada kesempatan kerja.

Menurut Badan Pusat Statistik, bekerja adalah kegiatan melakukan pekerjaan dengan tujuan memperoleh nafkah atau membantu memperoleh nafkah paling sedikit satu jam secara terus-menerus selama seminggu yang lalu. Sementara yang dimaksud dengan mencari pekerjaan adalah upaya yang dilakukan untuk memperoleh pekerjaan. Penduduk yang mencari pekerjaan dibagi menjadi penduduk yang pernah bekerja dan penduduk yang belum pernah bekerja.

Pengertian tenaga kerja dan bukan tenaga kerja hanya dibedakan oleh batasan umur yang masing-masing berbeda untuk setiap Negara. Di Indonesia batasan umur minimal 15 tahun tanpa batasan umur maksimal. Pemilihan batasan umur 15 tahun berdasarkan kenyataan bahwa pada umur tersebut sudah banyak

penduduk yang bekerja karena sulitnya ekonomi keluarga mereka. Indonesia tidak menganut batas umur maksimal karena Indonesia belum mempunyai jaminan sosial nasional. Hanya sebagian kecil penduduk Indonesia yang menerima tunjangan di hari tua yaitu pegawai negeri dan sebagian kecil pegawai perusahaan swasta. Untuk golongan inipun, pendapatan yang mereka terima tidak mencukupi kebutuhan mereka sehari-hari. Oleh sebab itu, mereka yang telah mencapai usia pensiun biasanya tetap masih harus bekerja sehingga mereka masih digolongkan sebagai tenaga kerja (Simanjuntak, 2005).

Menurut Badan Pusat Statistik seseorang dikatakan bekerja adalah :

1. Mereka yang selama seminggu sebelum pencacahan melakukan pekerjaan dengan maksud memperoleh atau membantu memperoleh penghasilan atau keuntungan yang lamanya bekerja paling sedikit satu jam dalam seminggu yang lalu.
2. Mereka yang selama seminggu sebelum pencacahan tidak melakukan pekerjaan atau bekerja kurang dari satu jam tapi mereka adalah:
 - a. Pekerja tetap, pegawai-pegawai pemerintah atau swasta yang sedang tidak termasuk kerja karena cuti, sakit, mogok, mangkir ataupun perusahaan menghentikan kegiatan sementara.
 - b. Petani yang mengusahakan tanah pertanian yang tidak bekerja karena menunggu hujan untuk menggarap sawah.
 - c. Orang-orang yang bekerja dibidang keahlian seperti dokter, dalang, dan lain-lain.

Sedangkan seseorang dikatakan tidak bekerja yaitu :

1. Mereka yang belum pernah bekerja, pada saat ini sedang berusaha mencari pekerjaan
2. Mereka yang sudah pernah bekerja, tapi pada saat pencacahan sedang menganggur dan berusaha mendapat pekerjaan.

2.2 Kesempatan Kerja

Kesempatan kerja mengandung pengertian bahwa besarnya kesediaan usaha produksi untuk mempekerjakan tenaga kerja yang dibutuhkan dalam proses produksi, yang dapat berarti lapangan pekerjaan atau kesempatan yang tersedia untuk bekerja yang ada dari suatu saat dari kegiatan ekonomi. Kesempatan kerja dapat tercipta apabila terjadi permintaan tenaga kerja di pasar kerja, sehingga dengan kata lain kesempatan kerja juga menunjukkan permintaan terhadap tenaga kerja (Budiriansyah,2017).

Kesempatan kerja berubah dari waktu ke waktu, perubahan tersebut terjadi akibat perubahan dalam perekonomian. Hal ini sesuai dengan konsep dalam ekonomi bahwa permintaan tenaga kerja merupakan permintaan turunan (derived demand) dari permintaan masyarakat terhadap barang dan jasa dalam perekonomian. Apabila perekonomian berkembang maka penyerapan tenaga kerja juga bertambah, pertumbuhan ekonomi mampu membawa pengaruh positif bagi kesempatan kerja dan produktivitas tenaga kerja (Simanjuntak, 2005).

Perluasan kesempatan kerja merupakan suatu usaha untuk mengembangkan sektor-sektor penampungan kesempatan kerja dengan produktivitas rendah. Usaha perluasan kesempatan kerja tidak terlepas dari faktor-faktor seperti, pertumbuhan jumlah penduduk dan angkatan kerja, pertumbuhan ekonomi, tingkat

produktivitas tenaga kerja, atau kebijaksanaan mengenai perluasan kesempatan kerja itu sendiri.

Kebijaksanaan negara dalam kesempatan kerja meliputi upaya-upaya untuk mendorong pertumbuhan dan perluasan kesempatan kerja di setiap daerah serta perkembangan kuantitas dan kualitas angkatan kerja yang tersedia agar dapat memanfaatkan seluruh potensi pembangunan di daerah masing- masing.

2.3 Faktor yang Mempengaruhi Partisipasi Tenaga Kerja

Partisipasi tenaga kerja dapat dipengaruhi oleh beberapa faktor. Berikut ini merupakan faktor yang mempengaruhi partisipasi tenaga kerja :

1. Usia

Simanjuntak (2001: 48) mengatakan bahwa semakin bertambah usia seseorang akan berpengaruh terhadap tingkat pekerjaan yang akan diperolehnya. Semakin dewasa seseorang maka keterampilan dalam bidang tertentu pada umumnya akan semakin meningkat, kekuatan fisik juga meningkat sehingga akan meningkatnya pekerjaan yang akan diterimanya. Pekerja yang bekerja di sektor formal yang banyak mengandalkan kemampuan fisik akan sangat terpengaruh oleh variabel umur.

Namun pada sisi lain, pada usia yang sudah tidak lagi produktif, keterampilan dan fisik seseorang akan mengalami penurunan. Hal ini sesuai dengan kenyataan bahwa dalam umur tersebut, banyak orang yang pensiun atau yang secara fisik sudah kurang mampu bekerja lagi, sehingga akan sangat berpengaruh terhadap tingkat pekerjaan yang akan diterima.

2. Kesehatan

Menurut Undang-Undang No.23 kesehatan adalah keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial dan ekonomi. Sedangkan menurut organisasi kesehatan dunia (WHO) pada tahun 1948 kesehatan adalah suatu keadaan fisik, mental, dan sosial kesejahteraan dan bukan hanya ketiadaan penyakit atau kelemahan. Kesehatan merupakan hal yang mutlak dibutuhkan oleh tubuh. Tanpa kesehatan, manusia tidak bisa beraktivitas dan bekerja. Karena apabila kesehatan buruk maka akan mempengaruhi konsentrasi dan fisik kita dalam bekerja sehingga hasil pekerjaan tidak akan maksimal

3. Jaminan Sosial

Menurut Undang-Undang No.3 tahun 1992 pasal 10 jaminan sosial tenaga kerja adalah suatu perlindungan bagi tenaga kerja dalam bentuk santunan berupa uang sebagai pengganti sebagian penghasilan yang hilang atau berkurang dan pelayanan sebagai akibat peristiwa atau keadaan yang dialami oleh tenaga kerja berupa kecelakaan kerja, sakit, hamil, bersalin, hari tua dan meninggal dunia.

Jaminan sosial tenaga kerja mempunyai beberapa aspek antara lain :

- a. Memberikan perlindungan dasar untuk memenuhi kebutuhan hidup minimal bagi tenaga kerja dan keluarganya.
- b. Merupakan penghargaan kepada tenaga kerja mendidik kemandirian pekerja sehingga pekerja tidak harus meminta belas kasiha orang lain jika dalam hubungan kerja terjadi resiko-resiko seperti kecelakaan kerja, sakit, hari tua dan lainnya.

4. Jenis Kelamin

Jumlah tenaga kerja laki-laki lebih besar dibandingkan perempuan karena ada kecenderungan laki-laki menjadi pencari nafkah utama dalam keluarga (Hartoko, 2018). Dalam pasar kerja tingkat partisipasi laki-laki masih dominan dibanding tingkat partisipasi perempuan. Perempuan biasanya terlibat dalam pekerjaan dengan produktivitas rendah. Beberapa penyebab rendahnya angka partisipasi perempuan dalam pasar kerja antara lain: 1) persepsi terkait peran domestik perempuan; 2) berkaitan dengan persepsi tersebut adalah perangkat pengukuran, penentuan, atau pendefinisian pekerjaan perempuan; 3) sifat musiman, paruh waktu, dan informal dari kebanyakan pekerjaan perempuan. (Kemenppa, 2016: 24).

5. Pendidikan

Tenaga kerja sebagai human capital (modal manusia) memegang peranan yang penting dalam sebuah perekonomian. Oleh karena itulah, investasi di bidang sumber daya manusia (SDM) perlu dilakukan. Investasi di bidang SDM dapat dilakukan melalui pendidikan dan latihan, migrasi, serta perbaikan gizi dan kesehatan.

Sumarsono (2009: 92) menjelaskan bahwa pendidikan dan latihan merupakan salah satu faktor penting dalam pengembangan SDM. Keduanya tidak hanya menambah pengetahuan, tetapi juga meningkatkan keterampilan dalam bekerja, kemandirian, maupun pembentukan kepribadian seorang individu sehingga meningkatkan produktivitas kerja. Menurut Undang-undang Republik

Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional Bab VI Pasal 14-19, jenjang pendidikan di Indonesia terdiri atas sebagai berikut:

- a. Pendidikan Dasar Pendidikan dasar di Indonesia diikuti oleh anak usia 7 sampai 15 tahun. Pendidikan dalam tahap ini dibagi menjadi dua tingkat yaitu Sekolah Dasar (SD/MI) serta Sekolah Menengah Pertama (SMP/MTs).
- b. Pendidikan Menengah Jenjang Pendidikan ini adalah lanjutan dari pendidikan dasar berupa Sekolah Menengah Atas atau Kejuruan (SMA/SMK) dan masing-masing membutuhkan waktu tempuh tiga tahun.
- c. Pendidikan Tinggi Setelah pendidikan dasar dan pendidikan menengah, jenjang pendidikan selanjutnya adalah pendidikan tinggi. Pendidikan tinggi kemudian dibagi lagi menjadi beberapa program pendidikan yaitu program diploma (D1, D2, D3), S1, S2 dan S3.

6. Status Dalam Rumah Tangga

Menurut Undang-Undang 52 tahun 2009 tentang Perkembangan Kependudukan dan Pembangunan Keluarga, Bab 1 Pasal 1 ayat 6 menjelaskan bahwa keluarga adalah unit terkecil dalam masyarakat yang terdiri dari suami istri atau suami, istri dan anaknya, atau ayah dan anaknya, atau ibu dan anaknya.

Status seorang individu sebagai anggota keluarga mempengaruhi kecenderungan partisipasi kerjanya. Anggota yang mempunyai peran cukup vital dalam keluarganya memiliki beban tanggungan yang lebih besar daripada anggota keluarga yang lain. Namun demikian, partisipasi kerja seorang individu dalam keluarga juga bergantung pada pola kerjasama antar anggota keluarga seperti apa yang diterapkan dalam keluarga tersebut. Penduduk lansia yang memiliki peran

sebagai kepala dalam keluarga memiliki pola partisipasi kerja yang tidak sama dibandingkan mereka yang berperan sebagai anggota keluarga lainnya.

2.4 Regresi Logistik Biner

2.4.1 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:

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

Untuk mempermudah penaksiran parameter dan untuk mendapatkan fungsi yang linier maka akan dilakukan transformasi pada persamaan (2.1), dengan menggunakan transformasi logit. Hasil transformasi diperoleh $g(x) = \beta_0 + \beta_1 x$ merupakan bentuk logit. Sedangkan untuk model regresi logistik biner dengan k variabel prediktor adalah sebagai berikut:

$$P(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + (\exp \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \quad (2.2)$$

dimana :

X_i : variabel independen, $i = 1, 2, \dots, k$

Y : variabel dependen,

β_0 : penduga parameter awal

$\beta_1, \beta_2, \dots, \beta_k$: penduga parameter 1,2,.....,k

k : banyaknya variabel independen

Jika respon Y yang mempunyai nilai 1 dengan peluang $P(Y = 1|X = x) = P(x)$ dan $g(x)$ adalah model regresi linear

$$g(x) = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

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

$$P(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2.3)$$

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

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

$$g(x) = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \quad (2.4)$$

$g(x)$ 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.

2.4.2 Uji Multikolinieritas

Uji Multikolinieritas bertujuan untuk mengetahui ada tidaknya hubungan antara variabel independen atau bebas. Pengujian ini dilakukan sebagai syarat digunakannya analisis berganda dimana regresi yang baik adalah regresi yang terbebas dari masalah multikolinieritas. (Ghozali, 2009).

Penentuan ada atau tidaknya multikolinieritas salah satunya dengan melihat nilai *Variance Inflation Factor* dan nilai *tolerance* dengan hipotesis sebagai berikut:

H_0 : tidak terdapat multikolinieritas

H_1 : terdapat multikolinieritas

Gejala multikolinearitas terjadi ketika $VIF \geq 10$ dan nilai $tolerance \leq 0,10$ (Ghozali, 2009).

2.4.3 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 memaksimalkan 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:

$$L(\beta) = \prod_{i=1}^n f(\beta, y) = \prod_{i=1}^n P(x_i)^{y_i} (1 - P(x_i))^{1-y_i} \quad (2.5)$$

Dengan :

$i = 1, 2, \dots, p$

$y_i =$ pengamatan pada peubah penjelas ke- i

$P(x_i) =$ peluang untuk peubah penjelas ke- i

Parameter β_i diduga dengan memaksimalkan persamaan diatas untuk mempermudah perhitungan maka dilakukan pendekatan logaritma sehingga fungsi log-likelihood sebagai berikut:

$$L(\beta) = \sum_{i=1}^n \{y_i \ln P(x_i) + (1 - y_i) \ln[1 - P(x_i)]\} \quad (2.6)$$

Nilai dugaan β_i dapat diperoleh dengan membuat turunan pertama dari $L(\beta)$ terhadap $\beta_i = 0$ dengan $i=1, 2, \dots, p$. Dari nilai tersebut dapat diketahui penduga dari

$P(x)$ dimana $g(x)$ adalah penduga logit sebagai fungsi linier dari peubah penjelas (Hosmer & Lemeshow 2000).

2.4.4 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 :

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

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

Dengan statistika uji

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

Kriteria uji : Tolak H_0 jika $G > \chi_{\alpha, k}^2$ atau nilai signifikansi kurang dari α .

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 $j = 1, 2, \dots, k$ (peubah X_j tidak berpengaruh nyata)

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

Dengan statistik uji :

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

Tolak H_0 jika $W > \chi_{\alpha,1}^2$ atau nilai signifikansi kurang dari α .

2.4.5 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 :

$$\chi^2 = \sum_{k=1}^g \frac{(O_k - n_k \pi_k)^2}{n_k \pi_k (1 - \pi_k)}$$

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 jika $X^2_{hitung} > X^2_{(db,\alpha)}$ atau nilai signifikansi kurang dari α .

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 n_1 dari grup 1 dan n_2 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}	$n_{1M} = n_1 - n_{1c}$
Grup 2 (n_2)	$n_{2M} = n_2 - n_{2c}$	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

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

2.5 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 populer 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. 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.

2. 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.

3. 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.

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{k_m} [S_{km}(X_{v(km)} - t_{km})] \quad (2.7)$$

dengan :

α_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 (2.7) model MARS dapat dituliskan sebagai berikut.

$$y_i = \alpha_0 + \sum_{m=1}^M \alpha_m \beta_{im}(x) + \varepsilon_i \quad (2.8)$$

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

$$y = \beta\alpha + \varepsilon \quad (2.9)$$

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 didefinisikan sebagai berikut.

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_m(x_i)]^2}{\left[1 - \frac{C(M)^2}{N}\right]} \quad (2.10)$$

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 = $\text{Trace} [B(B^T B)^{-1} B^T] + 1$

B : matriks basis fungsi

2.5.1 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 $f(x) = z = \text{logit } \pi(x)$. Model MARS untuk klasifikasi dapat dinyatakan sebagai berikut.

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

$$z = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [S_{km}(X_{v(km)} - t_{km})] \quad (2.11)$$

2.5.2 Ketepatan Klasifikasi

Ketepatan klasifikasi diperlukan untuk mengetahui pengelompokan data yang digolongkan dengan tepat pada kelompoknya. *Apparent Error Rate* (APER)

didefenisikan sebagai proporsi sampel yang tidak tepat diklasifikasikan (Johnson dan Wichern, 2007). Untuk mengetahui proporsi sampel yang tepat diklasifikasikan dapat dihitung dari nilai TAR (*Total Accuracy Rate*). Berikut ini merupakan tabel pengklasifikasian untuk respon biner.

Tabel 2.2 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.

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \times 100\% \quad (2.12)$$

dan

$$TAR(\%) = 1 - APER = 1 - \frac{n_{10} + n_{01}}{n} \times 100\% \quad (2.13)$$

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

2.6 Bootstrap Aggregating (Bagging)

Bootstrap Aggregating (Bagging) merupakan teknik yang diusulkan oleh Breiman (1996) yang dapat digunakan untuk me-reduksi variansi estimator pada metode klasifikasi dan regresi. Penggunaannya tidak dibatasi hanya untuk memperbaiki estimator. Teknik ini juga dapat memperbaiki stabilitas,

meningkatkan akurasi, dan kekuatan prediktif. Jumlah replikasi minimum untuk tujuan klasifikasi yakni sebanyak 50 kali, serta optimum saat nilai akurasi tertinggi telah didapatkan (Breiman, 1994). Bagging merupakan salah satu prosedur intensif perhitungan untuk memperbaiki estimator atau pengklasifikasi yang tidak stabil, khususnya masalah dimensi tinggi. Bagging merupakan salah satu bagian dari *Bootstrap*. Perbedaan pada Bagging dilakukan dengan membandingkan sampel secara berpasangan.

Buhlmann dan Yu (2002), secara singkat menyatakan algoritma Bagging sebagai berikut.

1. Sebuah data set L yang terdiri dari $\{(y_i, x_i), i = 1, 2, \dots, n\}$. Melakukan replikasi *bootstrap* pada data, sehingga didapatkan $L_i^* = (y_i^*, x_i^*)$, dengan $i = 1, 2, \dots, n$.
2. Replikasi *bootstrap* dilakukan sebanyak B kali, sehingga didapatkan $(L^{(B)})$ yang merupakan *resampling* pengembalian dari L .

Metode Bagging memiliki potensi untuk menurunkan kuadrat *error* peramalan pada berbagai proses, dan algoritma Bagging dalam pemodelan MARS adalah sebagai berikut (Buhlmann & Yu, 2002).

1. Terdapat data set L yang terdiri dari $\{(y_i, x_i), i = 1, 2, \dots, n\}$, dan melakukan replikasi *bootstrap* pada data, sehingga didapatkan $L_i^* = (y_i^*, x_i^*)$ atau disebut $(L^{(B)})$.
2. Melakukan pemodelan MARS pada $(L^{(B)})$.
3. Memprediksi variabel respon dari model MARS yang telah dihasilkan.
4. Mengulangi langkah 1 sampai dengan 3 hingga B replikasi *bootstrap*.

5. Melakukan prediksi pada variabel respon berdasarkan pemilihan prediksi yang sering muncul pada masing-masing observasi dari B replikasi *bootstrap* (maksimum *voting*).
6. Menghitung ketepatan klasifikasi prediksi model Bagging MARS.

