BAB II
TINJAUAN PUSTAKA

2.1 Konsep Kesejahteraan Petani

Rancangan Undang-Undang tentang Sistem Kesejahteraan Sosial Nasional (RUU SKSN) menyatakan bahwa kesejahteraan sosial adalah kondisi sosial ekonomi yang memungkinkan setiap warga negara untuk memenuhi kebutuhan yang bersifat jasmani, rohani dan sosial sesuai dengan harkat dan martabat manusia. Bertani merupakan salah satu mata pencaharian terbesar warga Negara Indonesia dan petani akan meningkat kesejahteraannya seiring dengan peningkatan kesejahteraan nasional Indonesia. Namun sensus pertanian tahun 2003 memberikan gambaran bahwa terdapat masalah serius dalam kemiskinan dan ketidaksejahteraan petani. Terjadi peningkatan 4.6 juta rumah tangga petani dari sensus Pertanian tahun 1993 ke 2003 atau dengan laju pertambahan 2.2% per tahun. Dari pertambahan tersebut jumlah petani *gurem* yaitu petani dengan luas lahan kurang dari 0.5 ha bertambah dari 10.8 juta atau 52.7% dari total rumah tangga petani pada tahun 1993 menjadi 13.7 juta pada tahun 2003 atau sekitar 56.5%. Pertambahan petani gurem ini mencapai 2.6% per tahun. Dengan demikian lebih banyak pertambahan petani *gurem* yang biasanya dianggap sebagai petani dengan kemiskinan tinggi (Tim Penerbit Buku Kompas, 2006).

Terdapat lima aspek yang dapat menunjukkan indikator kesejahteraan petani yaitu perkembangan struktur pendapatan, perkembangan pengeluaran untuk pangan, perkembangan Nilai Tukar Petani (NTP), perkembangan ketahanan pangan di tingkat rumah tangga petani dan daya beli rumah tangga petani (Sudana, et al., 2008 dalam Burhansyah dan Mella, 2010). Namun dalam penelitian ini indikator kesejahteraan ekonomi yang digunakan hanya pendapatan petani per bulan, jika petani memiliki rata-rata pendapatan per bulan di atas Upah Minimum Regional Kabupaten Kediri dan Blitar maka dikatakan bahwa petani itu memiliki kesejahteraan tinggi. Sebaliknya jika pendapatan petani di bawah Upah Minimum Regional maka petani tergolong berkesejahteraan rendah.

Onchan dan Chalamwong (Bainani, 2012) mengemukakan bahwa tiga sumber pendapatan rumah tangga petani berasal dari:

1. kegiatan farm atau usaha tani,
2. non-farm yaitu kegiatan perusahaan, pekerjaan upahan pada usahatani orang lain dan perusahaan pertanian, dan
3. kegiatan lain selain kegiatan pertanian (1 dan 2). Mereka mengemukakan juga bahwa semua aktivitas yang dilakukan di luar kegiatan usahatani sendiri yang dapat mendatangkan penghasilan disebut *off farm activities.*

2.2 Faktor-faktor yang Mempengaruhi Kesejahteraan Rumah Tangga Petani

Kesejahteraan sering dikaitkan dengan kemiskinan. Kemiskinan atau kesejahteraan suatu rumah tangga biasa dilihat dari keadaan sosial ekonomi rumah tangga tersebut. Indikator kesejahteraan rumah tangga yang digunakan pada penelitian ini adalah pendapatan. Menurut Kakishina (2010) peubah-peubah yang dapat mempengaruhi pendapatan usaha tani yang merupakan indikator kesejahteran keluarga petani adalah umur petani, tingkat pendidikan petani, jumlah anggota keluarga yang masih menjadi tanggungan dan luas lahan yang dimiliki.

Selain itu beberapa faktor tambahan yang juga mempengaruhi kesejahteraan rumah tangga menurut *AVRDC-The World Vegetable Centre* yang tercantum pada kuesioner BASELINE PROJEK USAID-INDONESIA adalah pekerjaan utama kepala rumah tangga yaitu petani atau non petani, pekerjaan sampingan kepala Rumah Tangga, status kepemilikan rumah, status kepemilikan lahan, sumber perolehan modal usaha, pemilikan kekayaan lain berupa bangunan lainnya, kepemilikan sumber informasi (televisi), alat komunikasi (telepon rumah atau telepon genggam), alat transportasi pribadi (sepeda motor dan mobil) dan banyaknya hewan ternak yang dimiliki(sapi dan kambing).

2.3 Statistika Deskriptif

Dalam mempelajari statistika, peneliti berkepentingan dengan penyajian dan penafsiran kejadian yang bersifat peluang yang terjadi dalam suatu penyelidikan terencana ataupun penelitian ilmiah. Walpole (1993) mendefinisikan bahwa metode statistika sebagai prosedur-prosedur yang digunakan dalam pengumpulan, penyajian, analisis dan penafsiran data. Walpole (1993) berpendapat bahwa

Statistika Deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistik deskriptif disajikan dalam bentuk tabel, diagram, grafik dan besaran-besaran lain (Walpole,1993).

2.4 Regresi Logistik

2.4.1 Model Regresi Logistik Biner

Analisis regresi logistik merupakan salah satu jenis analisis regresi yang digunakan jika peubah terikat (Y) bersifat kategorik dengan peubah penjelas (X) bersifat kategorik atau numerik. Jika peubah terikat terdiri dari dua kategori (*binary repons* atau *dichotomus repons)* maka model regresi logistik disebut sebagai regresi logistik biner.

Menurut Kutner, Nachtsheim dan Neter (2004), ketika peubah terikat merupakan peubah dikotomus atau terdiri dari dua ketegori (0 dan 1) maka peluang kejadian bagi Yadalah:

Y (Kejadian)	Peluang Kejadian
1(Sukses)	P(Y= 1) = π
0 (Gagal)	P(Y = 0) = 1-π

Hubungan antara k peubah penjelas $X_{1i}, X_{2i},...., X_{ki}$ dengan peluang peubah terikat π dianalisis dengan model peluang linier (Agresti, 2002):

$$\pi(X_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} \tag{2.1}$$

di mana:

n	= banyaknya pengamatan (i=1,2,…,n)
$\pi(Xi)$	= peluang terjadinya nilai peubah terikat Y pada peubah penjelas X ke- i
$X_{1i}, X_{2i},....,X_{ki}$	= peubah peubah penjelas ke-*i*
k	= banyaknya peubah penjelas
β_0	= intersep
β_j	= koefisien peubah penjelas ke - j

Persamaan 2.1 masih mengandung permasalahan, yaitu adanya kemungkinan menghasilkan nilai peubah terikat di luar interval [0,1], yang bertentangan dengan sifat peluang. Oleh karena itu, maka perlu dilakukan transformasi $\log_e$:

$$g(\pi) = \log_e\left(\frac{\pi}{1-\pi}\right) \tag{2.2}$$

Jika

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)} \quad ; \; i = 1,\ldots,n \tag{2.3}$$

maka fungsi logit dari model regresi logistik adalah (Hosmer dan Lemeshow, 2000):

$$\begin{aligned}
g(X_i) &= \ln\left(\frac{\pi(X_i)}{1-\pi(X_i)}\right) \\
&= \ln\left(\frac{\dfrac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki})}{1+\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki})}}{1-\dfrac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki})}{1+\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki})}}\right) \\
&= \ln\left(\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki})\right) \\
&= \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots. + \beta_k X_{ki} \\
&= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}
\end{aligned} \tag{2.4}$$

di mana:

n = banyaknya pengamatan

$\pi(x_i)$ = peluang sukses dengan syarat $X = x$ pengamatan ke-i

β_0, β_1 = koefisien regresi logistik

2.4.2 Pendugaan Parameter Regresi Logistik Biner

Nilai parameter model regresi logistik dapat diduga dengan beberapa metode, salah satunya adalah *Maximum Likelihood Estimation* (MLE). Secara umum hasil dugaan parameter pada metode ini diperoleh dengan cara memaksimumkan nilai peluang. Fungsi *likelihood* dalam metode ini merupakan fungsi peluang pengamatan untuk memperoleh penduga parameter yang tidak diketahui. Jika $\boldsymbol{\beta}$ adalah parameter model regresi logistik, maka metode *maximum likelihood* menduga nilai $\boldsymbol{\beta}$ dengan memaksimumkan fungsi Likelihood (Hosmer and Lemeshow, 2000).

$$L(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[1 - \pi(x_i)\right]^{1-y_i} \tag{2.5}$$

$$= \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[\frac{1 - \pi(x_i)}{\left(1 - \pi(x_i)\right)^{y_i}} \right]$$

$$= \prod_{i=1}^{n} 1 - \pi(x_i) \left[\frac{\pi(x_i)^{y_i}}{\left(1 - \pi(x_i)\right)^{y_i}} \right]$$

$$= \prod_{i=1}^{n} \left[1 - \pi(x_i)\right] \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right]^{y_i} \tag{2.6}$$

di mana $\pi(x_i) = \dfrac{\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)}$, (2.7)

maka $1 - \pi(x_i) = 1 - \dfrac{\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)}$

$$= \frac{1}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)} \tag{2.8}$$

sehingga

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{1}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)} \right] \left[\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right]^{y_i} \tag{2.9}$$

di mana: $j = 0, 1, \ldots, k$

$i = 1, 2, \ldots, n$

k = banyaknya peubah penjelas

n = banyaknya pengamatan

Persamaan 2.9 sulit diselesaikan, maka persamaan 2.9 diubah ke dalam bentuk log likelihood:

$$\ln L(\beta) = \ell(\beta)$$

$$\ell(\beta) = \ln \prod_{i=1}^{n} \left[\frac{1}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)} \right] \left[\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right]^{y_i}$$

$$= \ln \left[\frac{1}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)} \right]^{n} + \ln \left[\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right]^{\sum_{i=1}^{n} y_i}$$

$$= n \left[\ln 1 - \ln \left[1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right] \right] + \sum_{i=1}^{n} y_i \ln \left[\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right]$$

$$= \sum_{i=1}^{n} y_i \ln \left[\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right] - n \ln 1 - n \ln \left[1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right]$$

$$= \sum_{i=1}^{n} y_i \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) - n \ln \left[1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right] \qquad (2.10)$$

Untuk memaksimumkan persamaan log likelihood, maka dilakukan penurunan secara parsial terhadap β_j dan disamakan dengan nol.

$$\frac{\partial \ell(\beta)}{\partial(\beta_j)} = \frac{\partial \left(\sum_{i=1}^{n} y_i \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) - n \ln \left[1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right) \right] \right)}{\partial(\beta_j)} = 0$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n} y_i x_{ji} - n \sum_{i=1}^{n} x_{ji} \left[\frac{\exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}\right)} \right] = 0$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n} y_i x_{ji} - n \sum_{i=1}^{n} x_{ji} \pi(x_i) = 0 \qquad (2.11)$$

β_j yang diduga bersifat non linier maka penyelesaian persamaan (2.11) tidak dapat dilakukan secara langsung. Untuk mendapatkan penduga bagi parameter β_j diperlukan metode iterasi Newton Raphson.

Solusi yang diselesaikan dengan metode Newton Raphson untuk mendapatkan nilai $\hat{\boldsymbol{\beta}}$ yang memaksimumkan $L(\beta)$

Pandang vector **u**'$_{(1\times(k+1))}$ sebagai

$$\mathbf{u}' = \left(\frac{\partial L(\beta)}{\partial \beta_0}, \frac{\partial L(\beta)}{\partial \beta_1} .... \frac{\partial L(\beta)}{\partial \beta_p} \right) \tag{2.12}$$

Unsur vektor **u'**$_{(1\times(k+1))}$ didapatkan dari:

$$u_j^{(t)} = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \left(y_i - \pi(x_i)^{(t)} \right) x_{ij}$$

H adalah matriks Hessian dengan unsur-unsur

$$\mathbf{H}_{k+1} = \begin{bmatrix} h_{00} & \cdots & h_{0k} \\ \vdots & \ddots & \vdots \\ h_{\mathbf{k1}} & \cdots & h_{kk} \end{bmatrix}$$

di mana:

$$h_{ab}^{(t)} = \left. \frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b} \right|_{\beta^{(t)}} = -\sum_{i=1}^{n} x_{ia} x_{ib} \pi(x_i)^{(t)} \left(1 - \pi(x_i)^{(t)} \right) \tag{2.13}$$

Pandang Matriks $\mathbf{u}^{(t)}$ dan $\mathbf{H}^{(t)}$ sebagai matriks $\mathbf{u}$ dan $\mathbf{H}$ yang terevaluasi pada $\boldsymbol{\beta}^{(t)}$. $\boldsymbol{\beta}^{(t)}$ merupakan penduga $\boldsymbol{\beta}$ pada iterasi t (t=0,1,2,…). Nilai $\boldsymbol{\beta}^{(t)}$ pada setiap iterasi mengikuti bentuk deret Taylor orde kedua yaitu:

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\beta}^{(\mathbf{t})}) + \mathbf{u}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(\mathbf{t})}) + \tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(\mathbf{t})})' \mathbf{H}^{(t)} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(\mathbf{t})}) \tag{2.14}$$

dengan menyelesaikan $\dfrac{\partial L(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = \mathbf{u}^{(t)} - \mathbf{H}^{(t)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(\mathbf{t})}) = 0$

maka $\hat{\boldsymbol{\beta}}$ pada iterasi berikut adalah:

$$\boldsymbol{\beta}^{(\mathbf{t+1})} = \boldsymbol{\beta}^{(\mathbf{t})} - (\mathbf{H}^{(t)})^{-1} \mathbf{u}^{(t)} \tag{2.15}$$

$$\boldsymbol{\beta}^{(\mathbf{t+1})} = \boldsymbol{\beta}^{(\mathbf{t})} - \left\{ -\mathbf{X}' diag\left[\pi(\mathbf{x}_i^{(t)})(1 - \pi(\mathbf{x}_i^{(t)})) \right] \mathbf{X} \right\}^{-1} \mathbf{X}'(\mathbf{y} - \pi(\mathbf{x}_i^{(t)})) \tag{2.16}$$

Iterasi dihentikan jika $\pi(x_i)^{(t)}$ dan $\beta^{(t)}$ telah konvergen dan mendekati nilai duga dari $\hat{\boldsymbol{\beta}}$ untuk setiap j

$\left|\beta_j^{(t+1)} - \beta_j^{(t)}\right| \le c\left|\beta_j^{(t)} - \beta_j^{(t-1)}\right|^2$, c > 0 (Agresti,2002).

2.4.3 Pengujian Parameter

Pengujian parameter dilakukan untuk mengetahui apakah peubah penjelas berpengaruh secara simultan dan parsial terhadap peubah terikat.

a) Uji Parsial

Pengujian parameter secara parsial digunakan untuk mengetahui pengaruh setiap peubah penjelas terhadap peubah terikat. Pengujian ini dilandasi pada hipotesis (Agresti, 2002):

H_0 : $\beta_j = 0$

H_1 : $\beta_j \neq 0$

$\hat{\beta}_j \sim N(\beta_j, \sigma^2_{\beta_j})$

Jika H_0 benar, statistik uji Wald W_j,

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim Z \qquad (2.17)$$

di mana:

$$Var(\hat{\beta}_j) = \left[H(\hat{\beta}_j) \right]^{-1}$$

$$SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$$

Tolak H_0 jika $P\left[|Z| \ge W_j\right] < \alpha$.

b) Uji Simultan

Menurut Hosmer dan Lemeshow (2000), pengujian parameter secara simultan dilakukan untuk mengetahui apakah peubah-peubah penjelas secara bersama berpengaruh terhadap peubah terikat. Metode pengujian ini adalah membandingkan nilai pengamatan respon untuk model penuh dan model intersep berdasarkan statistik uji nisbah kemungkinan *(Likelihood Ratio Test)* *G* untuk hipotesis:

H_0 : $\beta_1 = \beta_2 = \ ... = \beta_p = 0$; $\beta_j = 0$

H_1 : Paling tidak terdapat satu j di mana $\beta_j \neq 0$ (*j =1, 2,…, p)*

Jika H_0 benar, maka Statistik uji nisbah kemungkinan adalah:

$$G = -2(L_0 - L_p) \sim \chi^2_{(p)} \quad (2.18)$$

di mana :

p = Banyaknya parameter

L_0 = log likelihood model regresi logistik tanpa peubah X_j

L_p = log likelihood model regresi logistik dengan peubah X_j

Tolak H_0 jika $P\left(\chi^2_{(p)} > G\right)$.

2.4.4 Uji Kesesuaian Model

Metode Hosmer dan Lemeshow (2000) adalah ukuran kesesuaian model. Statistik uji Hosmer-Lemeshow diperoleh dengan menerapkan uji khi kuadrat pada tabel kontingensi berukuran 2 × g. Tabel kontingensi merupakan hasil tabulasi silang antara peubah terikat biner dengan g kelompok peubah prediktor. Dalam praktek, digunakan 10 kelompok dan disebut sebagai ‘desil risiko’. Pengujian kesesuaian model menggunakan statistik Hosmer-Lemeshow berlandaskan hipotesis:

H_0 : Model sesuai (frekuensi pengamatan sama dengan frekuensi harapan)

H_1 : Model tidak sesuai (frekuensi pengamatan berbeda dengan frekuensi harapan)

Jika H_0 benar, maka statistik uji Hosmer and Lemeshow χ^2_{HL} :

$$\chi^2_{HL} = \sum_{k=1}^{g} \frac{(O_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \sim \chi^2_{(g-2)} \quad (2.19)$$

di mana

g = banyaknya kelompok (k=1,2,3….,g)

n_k = banyaknya individu pada grup ke-k

o_k = banyaknya nilai peubah terikat pada grup ke-k

$\bar{\pi}_k$ = merupakan rata-rata penduga peluang kejadian grup ke-k

Tolak H_0 jika $\chi^2_{HL} > \chi^2_{(\alpha,(g-2))}$ atau $P\left(\chi^2_{(g-2)} > \chi^2_{HL}\right)$. (Hosmer and Lemeshow, 2000).

2.4.5 Interpretasi Parameter Regresi Logistik

Interpretasi dari penduga slope koefisien regresi logistik b_j tidak semudah menginterpretasikan slope dari model regresi linier. Akibat dari penambahan 1 variasi unit X pada model regresi logistik sesuai dengan lokasi skala titik awal X sehingga diperlukan rasio odds sebagai interpretasi dari b_j. Interpretasi b_j dicari dalam persamaan model regresi logistik di mana penduga odds $\hat{\pi}/(1-\hat{\pi})$ meningkat secara multiplikatif oleh exp (b_j) untuk setiap unit peningkatan X.

Pandang nilai penduga peubah terikat regresi logistik pada $X = X_j$ sebagai

$$\hat{\pi}'(X_j) = b_0 + b_1 X_j \tag{2.20}$$

Notasi dari $\hat{\pi}'(X_j)$ mengindikasikan level X tertentu yang terkait dengan nilai penduga peubah terikat yang sesuai. Selanjutnya dipertimbangkan juga nilai penduga peubah terikat model logistik pada $X = X_j + 1$ sebagai:

$$\hat{\pi}'(X_j + 1) = b_0 + b_1(X_j + 1) \tag{2.21}$$

Selisih kedua persamaan tersebut dapat dituliskan sebagai berikut:

$$\hat{\pi}'(X_j + 1) - \hat{\pi}'(X_j) = b_1 \tag{2.22}$$

Berdasarkan persamaan(2.22), $\hat{\pi}'(X_j)$ adalah logaritma dari penduga odds ketika $X = X_j$, selanjutnya persamaan tersebut dilambangkan sebagai $\log_e(\text{odds}_1)$. Demikian juga persamaan $\hat{\pi}'(X_j + 1)$ merupakan penduga odds ketika $X = X_j + 1$ dan dilambangkan dengan $\log_e(\text{odds}_2)$. Oleh karena itu, selisih antara dua persamaan dapat juga ditulis sebagai berikut:

$$\log_e(odds_2) - \log_e(odds_1) = \log_e\left(\frac{odds_2}{odds_1}\right) = b_1 \tag{2.23}$$

Jika meng-antilog-kan kedua hasil persamaan tersebut, maka diperoleh rasio penduga dari odds dan dinamakan *odds ratio* dengan symbol $\widehat{OR}$ yang nilainya sama dengan $\exp(b_1)$:

$$\widehat{OR} = \frac{odds_2}{odds_1} = \exp(b_1) \tag{2.24}$$

(Kutner, et al., 2004)

2.4.6 Prosedur Klasifikasi

Pada model regresi logistik biner, penentuan kemampuan prosedur klasifikasi dalam memprediksi anggota kelompok peubah terikat biasa menggunakan peluang dari kesalahan klasifikasi yang disebut *error rate* atau *correct classification rate*. Penduga sederhana terhadap *Error Rate* diperoleh dengan melakukan prosedur klasifikasi dari himpunan data yang sama yang telah digunakan untuk menghitung fungsi klasifikasi. Metode ini sering disebut sebagai *resubstitution*. Setiap pengamatan y_{ij} yang disertakan pada fungsi klasifikasi akan menjadi anggota sebuah kelompok kategori peubah terikat tertentu (Y). Kemudian dihitung banyaknya pengamatan Y yang diklasifikasikan dengan tepat dan tidak tepat. Peluang kesalahan klasifikasi dari fungsi klasifikasi disebut dengan *apparent error rate,* yang ditampilkan dalam tabel klasifikasi regresi logistik biner berikut:

Tabel 2.1 Tabel Klasifikasi Regresi Logistik Biner

Y observasi	Y prediksi		Total
	Y_1	Y_2	
Y_1	n_{11}	n_{12}	n_1
Y_2	n_{21}	n_{22}	n_2

Banyaknya anggota pengamatan dari y_1 adalah n_1, n_{11} merupakan frekuensi anggota pengamatan yang terklasifikasi secara tepat pada y_1. Sebaliknya n_{12} tidak terklasifikasi secara tepat pada y_1, melainkan masuk pada anggota y_2. Besarnya n_1 merupakan penjumlahan dari n_{11} dan n_{12}. Sama halnya dengan y_1, frekuensi anggota y_2 dilambangkan dengan n_2, n_{21} merupakan anggota y_2 yang terklasifikasikan secara tepat. Banyaknya anggota y_2 yang tidak terklasifikasi secara tepat pada y_2 dilambangkan dengan n_{22} di mana $n_{21}+n_{22}=n_2$. Persamaan *apparent error rate* adalah:

$$Apparent\ Error\ Rate\ (\%) = \frac{n_{12}+n_{21}}{n_{11}+n_{12}+n_{21}+n_{22}} \tag{2.25}$$

(Rencher and Chistensen, 2012)

2.5 Bootstrap Aggregating (Bagging)

Bagging (*Bootstrap Aggregating*) diperkenalkan oleh Breiman (1994) adalah metode untuk memperbaiki kekuatan prediksi dari beberapa penduga atau algoritma tertentu seperti regresi atau pohon klasifikasi. Metode ini sangat baik digunakan pada data berdimensi tinggi di mana sering terdapat kesulitan dalam melakukan klasifikasi atau pemodelan. Prinsip kerja metode *bagging* adalah dengan cara mereduksi ragam dari peubah penjelas. Proses ini disebut sebagai *bagging* prediktor (Buhlmann, 2003).

Bagging prediktor merupakan metode untuk membangkitan beberapa macam peubah penjelas dan menggunakannya untuk mendapatkan suatu kumpulan peubah penjelas. Kumpulan peubah penjelas dibentuk dengan membuat pengulangan *bootstrap* dari sekumpulan data dan menggunakannya sebagai kumpulan data baru. Jika peubah penjelas bersifat numerik maka ukuran pemusatan data yang digunakan adalah rata-rata dan jika peubah penjelas bersifat kategorik maka ukuran pemusatan data berupa data yang sering muncul (modus).

Terdapat suatu kumpulan data yang terdiri dari $\{(y_n, x_n), n = 1......N\}$ di mana y bisa berskala numerik atau kategorik. Jika input berupa X maka Y diprediksi oleh $\varphi(x,\mathcal{L})$ dan $\varphi(x,\mathcal{L})$ adalah peubah penjelas. Misalkan terdapat gugus data $\{\mathcal{L}_k\}$ yang terdiri dari N pengamatan yang saling bebas dari sebaran yang sama. $\{\mathcal{L}_k\}$ merupakan pengulangan bootstrap yang digunakan untuk mendapatkan peubah penjelas yang lebih baik yang selanjutnya disebut $\{\varphi(x, \mathcal{L}_k)\}$.

Pengulangan *bootstrap* dilakukan sebanyak B kali dari $\mathcal{L}$ disebut dengan $\{\mathcal{L}^{(B)}\}$ sehingga dibentuk peubah penjelas $\{\varphi(x, \mathcal{L}^{(B)})\}$ di mana $\{\mathcal{L}^{(B)}\}$ adalah pengambilan contoh dengan pengembalian. Proses inilah yang disebut *bootstrap aggregating* dan disingkat dengan *bagging* (Breiman, 1994).

2.6 *Bagging* Regresi Logistik

Bagging telah banyak diaplikasikan di berbagai bidang, tetapi belum diaplikasikan pada bidang statistika seperti regresi logistik. *Bagging* regresi logistik bekerja dengan cara mengambil n bagian contoh dari data asli secara acak dengan pengembalian sebanyak B kali dan memodelkan regresi logistik untuk setiap bagian contoh pada setiap B pengulangan. Prediksi dari setiap pengamatan adalah rata-rata dari prediksi B (Perlich, Provost,and Simonoff, 2001)

Algoritma *bagging* regresi logistik biner menurut Efron dan Tibshirani (1993) dalam Ningrum (2012) adalah sebagai berikut :

1. Mengambil contoh *bootstrap* $\mathcal{L}^{(B)}$ berukuran n dari data $\mathcal{L}$ sebanyak B pengulangan dengan pengembalian.
2. Melakukan model regresi logistik biner pada data contoh *bootstrap* $\mathcal{L}^{(B)}$
3. Menghitung peluang peubah terikat untuk setiap pengamatan dan menghitung ketepatan klasifikasi. Kesalahan klasifikasi pada langkah ini disebut e_B.
4. Mengulang langkah 1 sampai langkah 3 sebanyak B kali (pengulangan *bootstrap*).
5. Memperoleh ketepatan klasifikasi *bagging* berdasarkan rata-rata ketepatan klasifikasi setiap pengambilan contoh sampai B, sehingga kesalahan klasifikasi *bagging* untuk B pengulangan adalah $\bar{e}_B$.
6. Membentuk model *bagging* regresi logistik biner dari rata-rata setiap parameter pada setiap pengambilan contoh sampai B.