

BAB I PENDAHULUAN

1.1 Latar Belakang

Regresi logistik biner merupakan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah peubah yang bersifat dikotomis (biner). Pada regresi logistik jika peubah respon terdiri dari dua kategori misalnya $Y = 1$ menyatakan hasil yang diperoleh “sukses” dan $Y = 0$ menyatakan hasil yang diperoleh “gagal” maka digunakan regresi logistik biner.

Dalam proses pemilihan model persamaan Regresi Logistik Biner “terbaik”, terdapat suatu tahapan yang sangat penting dan cukup menentukan, yaitu tahapan diagnostik model. Apabila di antara peubah prediktor ternyata terdapat korelasi yang cukup tinggi, maka terjadi multikolinieritas yang cukup tinggi (Retno, 2007). Gejala seperti multikolinieritas menimbulkan masalah dalam analisis. Menurut Gujarati (2003) jika multikolinieritas terjadi, koefisien regresi dari peubah prediktor tidak dapat ditentukan dan *standard error*-nya tidak terbatas.

Pada regresi logistik, asumsi tidak adanya multikolinieritas perlu diperhatikan. Sehingga beberapa pendekatan untuk mengatasi masalah tersebut dapat dicoba untuk mendapatkan kesimpulan yang lebih baik, misalnya PCLR_(s) (*Principle Component Logistic Regression Stepwise*) dan PLS-GLR (*Partial Least Square Generalized Linier Regression*).

Analisis dengan PCLR_(s) dan PLS-GLR menggunakan metode transformasi. PCLR_(s) menggunakan transformasi dari peubah prediktor asli ke peubah komponen utama, kemudian dipilih komponen utama berdasarkan metode *stepwise* sedangkan PLS-GLR menggunakan metode transformasi dari komponen PLS ke peubah aslinya. Maka, peneliti ingin mencoba suatu metode pemilihan model untuk mengatasi masalah multikolinieritas antara PCLR_(s) dan PLS-GLR.

Kedua metode tersebut akan diterapkan pada data sekunder yang di dapat dari MIS PT.Bank Rakyat Indonesia Tbk. KCB Malang yang menangani Kredit Usaha Mikro (KUM). Pada data tersebut, ingin

dimodelkan fungsi peluang pemberian kredit terhadap pengusaha mikro dari faktor sosial dan faktor finansial.

Peubah respon pada data tersebut bersifat kategori dikotomus, untuk itu model yang terbentuk adalah model logistik biner dan karena diantara peubah prediktornya mengandung sifat multikolinieritas maka akan di atasi dengan metode PCLR_(s) dan PLS-GLR yang nantinya dari kedua metode tersebut akan dipilih metode mana yang lebih baik dalam mengatasi multikolinieritas.

1.2 Perumusan Masalah

1. Bagaimanakah model regresi logistik biner yang terbentuk dari metode PCLR_(s) dan PLS-GLR dan apakah peubah *dept service ratio*, omzet penjualan, *current ratio*, biaya hidup, nilai agunan kredit, lama usaha dan jumlah tagunan mempengaruhi keputusan pemberian kredit pada pengusaha mikro?
2. Dari kedua metode, yaitu PCLR_(s) dan PLS-GLR manakah yang lebih baik digunakan dalam mengatasi multikolinieritas pada kasus data Kredit Usaha Mikro dilihat dari nilai AIC (*Akaike Information Criterion*), BIC (*Bayes Information Criterion*) dan PCP (*Percent Correct Prediction*)?

1.3 Batasan Masalah

Pada penelitian ini, permasalahan dibatasi pada data yang mengandung multikolinieritas pada sebagian peubah prediktor yang kontinyu dan peubah respon yang bersifat dikotomus atau biner.

1.4 Tujuan

1. Memilih metode terbaik dari data Kredit Usaha Mikro dalam mengatasi multikolinieritas pada variabel prediktor dengan variabel respon bersifat biner.
2. Mendapatkan model regresi logistik dengan tetap mempertahankan jumlah variabel prediktor yang ada dan mengetahui pengaruh variabel prediktor dari metode PCLR_(s) dan PLS-GLR pada data Kredit Usaha Mikro .

1.5 Manfaat Penelitian

Diperoleh informasi mengenai penanganan multikolinieritas pada data logistik dengan peubah respon bersifat dikotomis pada kasus data Kredit Usaha Mikro. Selain itu, dengan analogi yang sama metode penanganan multikolinieritas pada data-data lainnya yang peubah responnya bersifat kategori dikotomis dan peubah prediktornya saling berkorelasi.



UNIVERSITAS BRAWIJAYA



BAB II

TINJAUAN PUSTAKA

2.1 Regresi Logistik Biner

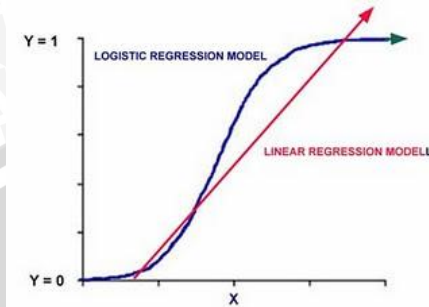
Analisis regresi adalah suatu analisis data yang mempelajari hubungan antara peubah respon (Y) dan satu atau lebih peubah prediktor (X), model yang sering digunakan adalah model regresi linier klasik dan metode pendugaan parameter menggunakan OLS (*Ordinary Least Square*). Pada regresi, peubah respon (Y) memiliki skala rasio, namun apabila peubah respon bersifat kategori metode analisis regresi linier menjadi tidak tepat digunakan, sehingga dikembangkan metode regresi logistik yang metode pendugaan parameter menurut Chatterjee dan Hadi (2006) menggunakan metode *Maximum Likelihood*.

Pendekatan model matematika yang dapat digunakan untuk menggambarkan hubungan dari beberapa peubah prediktor (X) dengan peubah respon (Y) yang bersifat kategorik adalah model Regresi Logistik. Pendekatan model lain mungkin saja digunakan, tapi regresi logistik sejauh ini merupakan model yang paling populer digunakan untuk menganalisis data jika peubah responnya bersifat kategori (Kleinbaum dan Klein, 2010).

Pada umumnya, peubah respon pada analisis regresi logistik bersifat kategori dengan hanya memiliki dua kemungkinan yang disebut dengan biner. Regresi logistik biner merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara peubah respon (Y) yang memiliki dua kemungkinan (sukses atau gagal) (Hosmer dan Lemeshow, 2000).

Dalam regresi logistik biner peubah respon bersifat dikotonomi yang mempunyai nilai 1 untuk kategori “sukses” atau 0 untuk kategori “gagal”, dan peubah prediktornya sebanyak k peubah, yaitu: $X_1, X_2, \dots, X_k$. Misalkan $Y = 1$ untuk kategori yang dinyatakan “sukses” dan $Y = 0$ untuk kategori yang dinyatakan “gagal” maka regresi logistik tersebut menggunakan regresi logistik biner.

Regresi logistik tidak memerlukan asumsi adanya hubungan antara peubah respon dan prediktor secara linier. Regresi logistik merupakan regresi non linier di mana model yang ditentukan akan mengikuti pola seperti gambar berikut :



Gambar 2.1 Kurva Regresi Logistik dan Regresi Linier

Menurut Agresti (2007) peubah Y yang bersifat biner lebih tepat dikatakan sebagai peubah indikator dan memenuhi sebaran Bernoulli. Jika banyaknya pengamatan lebih dari satu dan saling bebas maka y_i mengikuti sebaran binomial yaitu :

$$y_i \sim \text{Bin}(n_i, \pi_i)$$

$$y_i \sim \binom{n_i}{y} \pi_i^y (1 - \pi_i)^{n_i - y}, \quad n = 1, 2, \dots, n_i \quad (2.1)$$

Pada suatu penelitian, peubah peubah prediktor yang didapat adalah hasil pengukuran. Analisis yang dilakukan biasanya dengan mencari hubungan antara peubah respon dengan peubah prediktor model yang dibentuk adalah :

$$\pi_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.2)$$

Model pada persamaan (2.2) tersebut tidak baik digunakan kerana model tersebut memungkinkan penduga bertentangan dengan sifat peluang yaitu peubah respon diluar interval nol sampai dengan satu. Masalah tersebut dapat diatasi dengan transformasi :

$$g(x) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (2.3)$$

Dari persamaan (2.3) maka peubah respon berada di dalam interval nol sampai satu.

Untuk mempermudah pendugaan parameter regresi pada respon biner, maka $\pi(x)$ pada persamaan (2.2) di transformasikan menggunakan transformasi logit. Menurut Agresti (2002), bentuk model regresi logistik adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (2.4)$$

Proses transformasi tersebut adalah sebagai berikut :

$$\begin{aligned} \pi(x) &= \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \\ \{\pi(x)\} \{1 + \exp(\beta_0 + \beta_1 x)\} &= \exp(\beta_0 + \beta_1 x) \\ \{\pi(x)\} + \{\pi(x) \exp(\beta_0 + \beta_1 x)\} &= \exp(\beta_0 + \beta_1 x) \\ \pi(x) &= \{1 - \pi(x)\} \exp(\beta_0 + \beta_1 x) \\ \frac{\pi(x)}{1 - \pi(x)} &= \exp(\beta_0 + \beta_1 x) \\ \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) &= \ln\{\exp(\beta_0 + \beta_1 x)\} \\ \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) &= \beta_0 + \beta_1 x \\ g(x) &= \beta_0 + \beta_1 x \end{aligned} \quad (2.5)$$

$g(x)$ disebut bentuk logit. Sedangkan model regresi logistik dengan k peubah prediktor adalah :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \quad (2.6)$$

Jika model $\pi(x)$ di transformasikan menggunakan transformasi logit maka bentuk logit yang dihasilkan adalah :

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.7)$$

2.2 Pendugaan Parameter

Pendugaan parameter dari model logit menggunakan metode *maksimum likelihood*. Pada metode *maksimum likelihood* ingin diketahui nilai parameter mana yang mendekati nilai contoh sebenarnya (Hosmer dan Lemeshow, 2000).

Menurut Agresti (2002), metode *Maximum Likelihood* memberikan nilai pendugaan β untuk memaksimumkan dengan cara memaksimumkan fungsi likelihood dan mensyaratkan bahwa data harus mengikuti suatu sebaran tertentu. Pada regresi logistik, setiap pengamatan mengikuti sebaran Bernoulli.

Secara sistematis fungsi Likelihood untuk model regresi logistik dikotomis dapat ditulis sebagai berikut:

$$f(\beta, y) = \pi(x)^y (1 - \pi)^{1-y} \quad (2.8)$$

Likelihood amatan merupakan perkalian dari masing-masing fungsi Likelihood:

$$\xi(\beta) = \prod_{i=1}^n f(\beta, y) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi)^{1-y_i} \quad (2.9)$$

Secara matematis akan lebih mudah untuk memaksimumkan $\log \xi(\beta)$ atau disebut juga log Likelihood yang dinotasikan sebagai $L(\beta)$, yakni dengan cara mendiferensialkan $L(\beta)$ terhadap β dan menyamakannya dengan nol.

$$\begin{aligned} L(\beta) &= \xi(\beta) \\ &= \sum_j [\sum_i y_i x_{ij}] \beta_j - \sum_i n_i \log [1 + \exp \beta_j x_{ij}] \\ \frac{\partial L}{\partial \beta_a} &= \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \left[\frac{\exp \beta_j x_{ij}}{1 + \exp \beta_j x_{ij}} \right] \\ 0 &= \sum_i y_i x_{ia} - \sum_i n_i \hat{\pi}_i x_{ia} \quad a=0, \dots, k \end{aligned} \quad (2.10)$$

$$\text{di mana } \pi(x_i) = \frac{\exp(\sum_j^k \beta_j x_{ij})}{1 + \exp(\sum_j^k \beta_j x_{ij})} \quad (2.11)$$

Agar nilai fungsi $L(\beta_j)$ mencapai maksimum maka turunan parsial pertama terhadap β_j disama dengarkan nol.

$$\begin{aligned}\frac{\partial L(\beta_j)}{\partial \beta_j} &= \sum_{j=0}^p [\sum_{i=1}^n y_i x_{ij}] \beta_{ij} - n \ln \left[\exp \left[\sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} \right] \right] \\ \frac{\partial L(\beta_j)}{\partial \beta_j} &= \sum_{j=0}^p [\sum_{i=1}^n y_i x_{ij}] \beta_{ij} \sum_{i=1}^n n_i x_{ij} \left[\frac{\exp \left(\sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} \right)}{1 + \exp \left(\sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} \right)} \right] \\ &= \sum_{i=1}^n \sum_{j=0}^p y_i x_{ij} - \sum_{i=1}^n n_i x_{ij} \left[\frac{\exp \left(\sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} \right)}{1 + \exp \left(\sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} \right)} \right] = 0\end{aligned}\quad (2.12)$$

$$= \sum_{i=1}^n \sum_{j=0}^p y_i x_{ij} - \sum_{i=1}^n n_i x_{ij} \pi_i = 0 \quad (2.13)$$

Untuk regresi logistik yang menjelaskan persamaan (2.11) adalah non linier terhadap β_j sehingga membutuhkan metode iterasi yang khusus untuk pengerjaan dimana dalam perhitungannya dibantu oleh komputer. Adapun metode iterasi yang digunakan yaitu iterasi *Newton-Raphson*. Metode ini memperoleh dugaan maksimum likelihood bagi β dengan iterasi yang menggunakan rumus (Agresti, 2007) :

$$\beta_{(m+1)} = \beta_{(m)} - h_{(ab)}^{-1} g_{(m)}$$

m = 1,2,3,... sampai konvergen dengan nilai matriks

$$h_{ab}^{(t)} = \left. \frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b} \right|_{\beta^t} = - \sum_i x_{ia} x_{ib} n_i \pi_i^{(t)} (1 - \pi_i^{(t)}) \quad (2.14)$$

$$u_j^t = \left. \frac{\partial L(\beta)}{\partial \beta_j} \right|_{\beta^{(t)}} = \sum_i (y_i - n_i \pi_i^{(t)}) x_{ij} \quad (2.15)$$

Dimana $u = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}$ adalah vektor gradien dan $\begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{k1} & \cdots & h_{kk} \end{bmatrix}$ matrik

hessian. Iterasi untuk memperoleh $\beta_{(m)}$ terus dilakukan sampai mencapai konvergen c untuk setiap j, yaitu :

$$|\beta_j^{(m+1)} - \beta_j| \leq c |\beta_j^{(m)} - \beta_j|^2 \text{ untuk } c > 0$$

2.3 Pengujian Signifikansi Parameter

Setelah menduga parameter maka langkah selanjutnya yang dilakukan adalah menguji signifikansi parameter tersebut. Untuk itu digunakan uji hipotesis statistik untuk menentukan apakah peubah prediktor dalam model berpengaruh nyata terhadap peubah respon. Pengujian signifikansi parameter dilakukan sebagai berikut

- a. Pengujian signifikansi parameter ($\hat{\beta}$) secara parsial

Uji koefisien regresi secara parsial digunakan untuk memeriksa peranan koefisien regresi dari masing-masing peubah prediktor secara individu dalam model.

Hipotesis:

H_0 : $\beta_j = 0$ dengan β_j adalah parameter ke – j

H_1 : $\beta_j \neq 0, j = 1, 2, \dots, n$

Statistik uji :

$$w = \frac{\beta_j}{se(\beta_j)} \quad (2.16)$$

Rasio yang dihasilkan dari uji *wald* adalah , dibawah hipotesis H_0 , akan mengikuti sebaran normal baku. Kriteria penolakan H_0 jika nilai $|W| > Z_{\alpha/2}$ (Hosmer dan Lemeshow, 2000).

- b. Pengujian signifikansi parameter (β) secara simultan

Pegujian signifikansi parameter secara simultan dilakukan sebagai upaya memeriksa peranan peubah prediktor dalam model secara bersama-sama.

Hipotesis :

$H_0 : \beta = \beta_1 = \dots = \beta_k = 0$

H_1 : Minimal terdapat satu j di mana $\beta_j \neq 0, j = 1, 2, \dots, k$

Uji yang digunakan adalah uji nisbah kemungkinan (*Likelihood Ratio Test*) yaitu:

$$G^2 = \left[\frac{L_1}{L_0} \right]$$

$$G^2 = -2\ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} - \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i(1-\hat{\pi}_i)^{(1-y_i)}}} \right] \quad (2.17)$$

dengan:

- n_1 : banyaknya amatan yang berkategori 1
 - n_0 : banyaknya amatan yang berkategori 0
 - L_1 : *likelihood* tanpa prediktor tertentu
 - L_0 : *likelihood* dengan prediktor tertentu
- (Kleinbaum dan Klein, 2010)

Jika H_0 benar maka G berdistribusi $\chi^2_{(v)}$, dengan v adalah selisih banyaknya parameter yang diduga pada kedua model. Hipotesis nol akan ditolak jika nilai $P[\chi^2_{(v)} > G] < \alpha$.

Hal ini mengindikasikan bahwa paling sedikit ada satu β_j yang tidak sama dengan nol.

Walaupun uji Wald cukup memadai untuk contoh yang besar, uji Likelihood lebih akurat untuk ukuran contoh yang sering digunakan pada penelitian (Agresti, 2007).

2.4 Interpretasi Koefisien Regresi Logistik

Proses selanjutnya adalah interpretasi terhadap koefisien parameter tersebut. Interpretasi dalam regresi logistik menggunakan nilai odds ratio yang menunjukkan perbandingan tingkat kecenderungan dari kategori yang ada dalam satu peubah prediktor.

Nilai odds ratio dari regresi logistik didefinisikan sebagai berikut:

$$\psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

$$\psi = e^{\beta_1} \quad (2.18)$$

Sedangkan nilai log rasio adalah :

$$\ln \psi = \ln \left[\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} \right] = \ln \left[\frac{\pi(1)}{[1-\pi(1)]} \right] - \ln \left[\frac{\pi(0)}{[1-\pi(0)]} \right] = -g(0)$$

$$\text{Ln}(\psi) = \ln(e^{\beta_1}) = \beta_1 \quad (2.19)$$

Nilai Odds Ratio ψ digunakan untuk menunjukkan kecenderungan hubungan suatu peubah X terhadap peubah Y dan selalu bernilai positif. Misalnya $\psi = 1$ berarti bahwa $x = a$ mempunyai resiko yang sama dengan $x = b$ untuk menghasilkan $y = j$. Apabila $1 < \psi < \infty$ berarti $x = a$ memiliki resiko lebih tinggi ψ kali dibandingkan $x = b$ untuk menghasilkan $y = j$ dan sebaliknya.

2.5 Uji kesesuaian model (*Goodness of fit*)

Untuk menguji apakah model yang dihasilkan sudah sesuai atau tidak, maka perlu dilakukan pengujian kesesuaian model atau *goodness of fit*. Hipotesis yang digunakan dalam pengujian ini yaitu:

H_0 : Model sesuai vs H_1 : Model tidak sesuai

Adapun statistik uji yang digunakan untuk menguji kesesuaian model adalah uji *Pearson* dan uji *Deviance*.

1. Statistik *Pearson*

$$\chi^2_{\text{Pearson}} = \sum_{a=1}^h \chi_p^2(y_a, \hat{\pi}_a) \quad (2.20)$$

$$\text{di mana } \chi_p^2(y_a, \hat{\pi}_a) = \sum_{b=1}^g \frac{(y_{ab} - n_a \hat{\pi}_{ab})^2}{n_a \hat{\pi}_{ab}}$$

2. Statistik *Deviance*.

$$D = 2 \sum_{a=1}^h \chi_D^2(y_a, \hat{\pi}_a) \quad (2.21)$$

$$\text{di mana } \chi_D^2(y_a, \hat{\pi}_a) = \sum_{b=1}^g y_{ab} \ln \left[\frac{y_{ab}}{n_a \hat{\pi}_{ab}} \right]$$

Dengan :

$\hat{\pi}_{ab}$: peluang Y pada kategori ke-b dan X ke-a.

y_{ab} : pengamatan Y pada kategori ke-b dan X ke-a.

p : banyaknya parameter pada model.

h : banyaknya peubah prediktor X.

g : banyaknya kategori peubah respon Y.

Statistik *Pearson* dan *deviance* menyebar mengikuti sebaran khi kuadrat dengan derajat bebas v . Keputusan menolak H_0 jika $\chi^2_{pearson} > \chi^2_{(v)}$ dan $D > \chi^2_{(v)}$ atau $P[\chi^2_{(v)} > \chi^2_{pearson}]$ dan $P[\chi^2_{(v)} > D]$ lebih kecil dari peluang yang diinginkan (α), sehingga dapat diambil kesimpulan bahwa model yang diperoleh tidak sesuai (Agresti,2002).

2.6 Analisis Komponen Utama (*Principal Component Analysis*)

Analisis komponen utama merupakan suatu struktur yang menerangkan ragam peragam melalui kombinasi linier dari peubah-peubah dengan ragam maksimum (Escabias, dkk.,2005).

Beberapa hal berikut dapat digunakan untuk menentukan komponen utama yang akan digunakan, yaitu menerima seluruh komponen utama, menerima komponen utama dengan akar ciri terbesar untuk meminimalkan keragaman peubah prediktor, atau menerima komponen utama yang berkorelasi tinggi dengan peubah respon untuk meminimalkan kuadrat tengah galat.

Pada matriks $\mathbf{X}$ dengan p peubah kontinyu dan n ukuran contoh di mana elemen dari matriks tersebut dapat dituliskan $\mathbf{X} = [x_{ij}]$ ($i=1,2,\dots,n$; $j=1,2,\dots,p$), vektor kolom dari matriks tersebut adalah $X_1, X_2, \dots, X_p$.

Matriks ragam peragam dari matriks tersebut dapat dinyatakan sebagai berikut:

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^T (x_{ij} - \bar{x}_j) \quad , \quad \text{di mana rata-rata dari contoh}$$
$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (j = 1,2,\dots,p) \quad (\text{Escabias,dkk.,2005}) \quad (2.22)$$

Pada analisis komponen utama, vektor peubah asal, yaitu

$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)$ ditransformasi menjadi vektor peubah baru $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q)$ di mana $q=p$, dalam bentuk persamaan sebagai berikut:

$\mathbf{W} = \mathbf{V} \mathbf{X}$, di mana $\mathbf{V}$ adalah matriks dari vektor ciri dan elemen dari matriks tersebut dapat dituliskan $\mathbf{V} = \mathbf{v}_{jk}$ ($j=1,2,\dots,p$; $k=1,2,\dots,p$). Ragam dari $\mathbf{W}$ dinyatakan sebagai $\text{Var}(\mathbf{W}) = \mathbf{V}^T \mathbf{S} \mathbf{V}$. Agar ragam $\mathbf{W}$ maksimum, digunakan batasan $\mathbf{V}^T \mathbf{V} = 1$ dan dengan menggunakan metode pengganda Lagrange didapatkan:

$$U(\mathbf{V}, \lambda) = \mathbf{V}^T \mathbf{S} \mathbf{V} - \lambda (\mathbf{V}^T \mathbf{V} - 1) \quad (2.23)$$

Fungsi ini akan maksimum jika turunan parsial pertama $U(\mathbf{V}, \lambda)$ terhadap $\mathbf{V}$ dan λ sama dengan nol, sehingga:

$$\begin{aligned} \frac{\partial U(\mathbf{V}, \lambda)}{\partial \lambda} &= \mathbf{V}^T \mathbf{V} - 1 = 0 \\ &= \mathbf{V}^T \mathbf{V} = 1 \end{aligned}$$

$$\frac{\partial U(\mathbf{V}, \lambda)}{\partial \mathbf{V}} = 2 \mathbf{S} \mathbf{V} - 2 \lambda \mathbf{V} \quad (2.24)$$

$$2 \mathbf{S} \mathbf{V} - 2 \lambda \mathbf{V} = 0$$

$$2(\mathbf{S} - \lambda \mathbf{I}) \mathbf{V} = 0$$

$$(\mathbf{S} - \lambda \mathbf{I}) \mathbf{V} = 0 \quad (2.25)$$

Untuk mendapatkan penyelesaian dari persamaan (2.23) yang memenuhi syarat $\mathbf{V}^T \mathbf{V} = 1$, maka $\mathbf{S} - \lambda \mathbf{I}$ harus bersifat singular atau harus memenuhi persamaan:

$$|\mathbf{S} - \lambda \mathbf{I}| = 0 \quad (2.26)$$

(Neil, 2002),

Persamaan (2.24) merupakan persamaan karakteristik dari matriks ragam peragam, sehingga diperoleh akar-akar karakteristik $\lambda_1, \lambda_2, \dots, \lambda_p$ di mana $\lambda_1 \geq \dots \geq \lambda_p \geq 0$. Kemudian dari setiap akar ciri diperoleh vektor ciri yang didapat dari :

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1k} \\ v_{21} & v_{22} & \dots & v_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{j1} & v_{j2} & \dots & v_{jk} \end{bmatrix}$$

V digunakan untuk mentransformasi peubah X ke dalam skor komponen utama.

Jika persamaan (2.20) dikalikan dengan V^T dan $V^T V = 1$, maka akan diperoleh hasil:

$$2 V^T S V - 2 \lambda = 0 \longrightarrow \lambda = V^T S V$$

Sehingga ragam dari W adalah :

$$\text{Var} (W) = \text{Var} (V X) = V^T S V = \lambda$$

Keragaman yang dapat dijelaskan oleh komponen utama ke- i terhadap keragaman total adalah:

$$\frac{\lambda_i}{\sum_{j=1}^p \lambda_j} \times 100\% \quad (2.27)$$

Sedangkan keragaman total yang dapat dijelaskan oleh q komponen utama pertama adalah:

$$\frac{\sum_{i=1}^q \lambda_i}{\sum_{j=1}^p \lambda_j} \times 100\% \quad (2.28)$$

di mana q adalah banyaknya komponen utama yang dimasukkan dalam model. Semakin besar nilai ukuran kesesuaian tersebut, maka makin layak q komponen utama pertama tersebut digunakan

Menurut Neil (2002) menyatakan, apabila komponen utama telah diperoleh, maka tahap selanjutnya adalah menentukan skor komponen utama. Skor komponen utama dari setiap individu akan digunakan untuk analisis lanjut. Jika nilai pengamatan dari individu ke- i pada peubah asal ke- j adalah $X_{i1}, X_{i2}, \dots, X_{ij}$, maka skor komponen dari individu ke- i untuk komponen utama W_i yang dihasilkan dari matriks ragam peragam adalah :

$$SK_{ij} = [v_{1j}, v_{2j}, \dots, v_{kj}] \begin{bmatrix} X_{i1} - \bar{X}_1 \\ X_{i2} - \bar{X}_2 \\ \dots \\ X_{ip} - \bar{X}_p \end{bmatrix} \quad (2.29)$$

$$\text{atau } SK_{ij} = \mathbf{V}_j^T (\mathbf{X}_i - \bar{\mathbf{X}})$$

Jika komponen utama dihasilkan dari matriks korelasi, maka matriks data individu ($\mathbf{X}$) digantikan oleh matriks data skor baku $\mathbf{X}^*$, di mana :

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{\text{Var}(x_j)}} \quad (2.30)$$

Prosedur penentuan jumlah komponen utama yang digunakan dalam membentuk model:

1. Ambil akar ciri yang lebih besar dari 1 ($\lambda_i > 1$)
2. Pilih q buah komponen utama sebagai penyumbang terbesar terhadap total keragaman lebih besar dari 75%.

2.7 Metode Penanganan Multikolinieritas pada Regresi Logistik Biner

Apabila di antara peubah prediktor ternyata terdapat korelasi cukup tinggi, maka akan terjadi multikolinieritas. Sehingga perlu dilakukan penanganan untuk mengatasi multikolinieritas. Terdapat beberapa metode untuk mengatasi multikolinieritas yaitu metode analisis komponen utama dengan *stepwise* (PCLR_(s)) dan PLS-GLR (Hosmer dan Lemeshow, 2000).

2.7.1 *Principal Component Logistic Regression Stepwise* (PCLR_(s)).

Hosmer dan Lemeshow (2000) menyatakan bahwa metode *stepwise* sangat penting dalam pemeriksaan peubah prediktor di dalam suatu model. *Stepwise* dalam model regresi logistik urutan penyisipannya ditentukan dengan menggunakan nilai taraf nyata P_E dan P_R yang dibandingkan dengan nilai *probability* (*Likelihood Ratio Statistic*) apakah suatu peubah prediktor tersebut tetap berada dalam model atau tidak. Taraf nyata P_E digunakan untuk mengontrol apakah suatu peubah prediktor dimasukkan ke dalam model, sedangkan taraf nyata P_R digunakan untuk mengontrol apakah suatu peubah prediktor tetap berada dalam model atau dikeluarkan dari model. Penentuan taraf nyata P_E dan P_R yang direkomendasikan adalah 0.15 untuk P_E dan 0.20 untuk P_R (Hosmer dan Lemeshow, 2000). Untuk itu akan digunakan

metode *stepwise* yang akan dikombinasikan dengan metode komponen utama dalam mengatasi multikolinieritas.

Prosedur awal dari metode ini yaitu melakukan tranformasi peubah prediktor asli ke peubah komponen utama sehingga, $\hat{\gamma}_{(s)} = (\hat{\gamma}_{0(s)}, \hat{\gamma}_{1(s)}, \dots, \hat{\gamma}_{s(s)}) \neq (\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_s)$.

Dalam pemilihan 's' komponen utama digunakan metode *stepwise*, yaitu dengan memasukkan komponen utama secara bertahap kemudian menyeleksi komponen tersebut secara bertahap pula.

Dalam bukunya (Hosmer dan Lemeshow, 2000) menjelaskan langkah-langkah penentuan model terbaik dengan metode *stepwise* adalah:

1. Menentukan P_E dan P_R untuk menyeleksi peubah-peubah yang masuk ke dalam model maupun yang keluar, di mana P_E adalah taraf nyata untuk mengontrol apakah suatu peubah prediktor dimasukkan ke dalam model, sedangkan P_R adalah taraf nyata untuk mengontrol apakah suatu peubah prediktor tetap berada dalam model.
2. langkah (0) menganggap model memiliki p peubah prediktor yang semuanya dianggap penting sebagai peubah yang berpengaruh dalam model.
 - a. langkah ini dimulai dengan membetuk model yang hanya mengandung intersep saja kemudian menghitung nilai *log-likelihood* (L_0).
 - b. Kemudian hitung nilai *log-likelihood* untuk model yang mengandung peubah W_j dan dituliskan sebagai $L_j^{(0)}$, di mana j adalah indeks untuk peubah yang ditambahkan ke dalam model dan (0) adalah indeks untuk langkah.
 - c. Hitung nilai nisbah kemungkinan untuk model yang mengandung peubah W_j melawan model yang hanya mengandung intersep saja, $G_j^{(0)} = -2(L_0 - L_j^{(0)})$ dan nilai p yaitu $P_j^{(0)}$, $P_j^{(0)} = P(\chi_{(v)}^2 > G_j^{(0)})$.
 - d. Bandingkan nilai $P_j^{(0)}$, peubah yang paling penting adalah peubah yang memiliki nilai $P_j^{(0)}$ terkecil. Jika peubah tersebut dituliskan sebagai W_{e1} maka $P_{e1}^{(0)} = \min(P_j^{(0)})$. Indeks "ei"

digunakan untuk menotasikan peubah yang menjadi calon peubah yang masuk pada langkah (1). Peubah tersebut cukup penting untuk masuk ke dalam model jika $P_{e1}^{(0)} < P_E$, jika tidak maka proses berhenti.

3. Langkah (1), dimulai dengan model regresi logistik yang mengandung W_{e1}
 - a. Hitung $L_{e1}^{(1)}$, yaitu *log-likelihood* model.
 - b. Tentukan apakah sisa $p-1$ peubah adalah penting sekali untuk model ketika peubah W_{e1} berada dalam model, dengan membuat $p-1$ model regresi logistik yang mengandung X_{e1} dan X_j , $j = 1, 2, 3, \dots, p-1$ dan $j \neq e1$. Hitung nilai *log-likelihood* dan dituliskan sebagai $L_{e1j}^{(1)}$.
 - c. Hitung nilai nisbah kemungkinan ini melawan model yang hanya mengandung W_{e1} ($G_j^{(1)} = -2(L_{e1}^{(1)} - L_{e1j}^{(1)})$), hitung nilai p statistik ini sebagai $P_j^{(1)}$.
 - d. Misalkan peubah dengan nilai $P_j^{(1)}$ terkecil pada langkah (1) W_{e2} di mana $P_{e2}^{(1)} = \min(P_j^{(1)})$. Jika $P_{e2}^{(1)} < P_E$ lanjutkan langkah (2), jika tidak maka proses berhenti.
4. Langkah (2), dimulai dengan menduga model regresi logistik yang mengandung W_{e1} dan W_{e2} . Secara umum langkah ini adalah memilih model yang sesuai dengan menghapus peubah yang ditambahkan pada tahap sebelumnya dan menetapkan peubah penting selanjutnya yang dimasukkan.
 - Seleksi langkah mundur.
 - a. Hitung $L_{-ej}^{(2)}$ yaitu *log-likelihood* model dengan W_{ej} yang diabaikan.
 - b. Hitung nilai nisbah kemungkinan model ini melawan model penuh pada langkah (2) yaitu $G_{-ej}^2 = -2(L_{ele2}^{(2)} - L_{-ej}^{(2)})$ dan $P_{-ej}^{(2)}$ sebagai nilai p .
 - c. Untuk memastikan apakah suatu peubah harus dihapus dari model, pilih peubah yang menghasilkan $P_{-ej}^{(2)}$ maksimum, yang dituliskan sebagai W_{r2} , $P_{r2}^{(2)} = \max(P_{-e1}^{(2)}, P_{-e2}^{(2)})$.

- d. Untuk memutuskan apakah X_{r2} harus dibuang, bandingkan $P_{r2}^{(2)}$ dengan P_R . Jika $P_{r2}^{(2)} > P_R$ maka X_{r2} keluar dari model. Jika $P_{r2}^{(2)} < P_R$ maka X_{r2} tetap berada dalam model.

Seleksi langkah maju dengan cara :

- a. Pada tahap seleksi selanjutnya, $p-2$ model regresi logistik mengandung W_{e1} , W_{e2} dan W_{ej} , untuk $j = 1, 2, 3, \dots, p, j \neq e1, e2$. Hitung nilai nisbah kemungkinan model ini melawan model yang hanya mengandung W_{e1} dan W_{e2} dan tentukan nilai $P_j^{(2)}$.
 - b. Dapatkan W_{ej} peubah yang memiliki nilai $P_j^{(2)}$ minimum, $P_{e3}^{(2)} = \min(P_j^{(2)})$. Jika $P_{e3}^{(2)} < P_E$ lanjutkan ke langkah (3), jika tidak proses berhenti.
5. Langkah (3), prosedur pada langkah (3) sama dengan prosedur langkah (2). Langkah terus berlanjut sampai langkah terakhir atau langkah (S).
6. Langkah (S), ini terjadi ketika : (1) semua peubah bebas masuk dalam model atau (2) semua peubah yang berada dalam model mempunyai nilai p untuk dipindahkan kurang dari P_R , dan peubah yang belum masuk dalam model mempunyai nilai p untuk masuk dalam model melebihi P_E .

2.7.2 PLS-GLR (*Partial Least Square Generalized Linier Regression*)

Metode PLS adalah metode yang dapat digunakan untuk memprediksi peubah respon dari sekelompok besar peubah prediktor (Abdi, 2003). Metode PLS melibatkan peubah respon dalam pembentukan komponen laten. Metode PLS mengidentifikasi kombinasi linier dari peubah asal sebagai peubah prediktor dan menggunakan komponen linier ini dalam regresi linier.

Menurut Jian,dkk.(2006) metode PLS-GLR mereduksi peubah menjadi peubah atau komponen baru dengan banyak komponen dengan jumlah kurang dari atau sama dengan dimensi terkecil matriks peubah prediktor. Komponen baru yang telah terbentuk merupakan komponen yang saling bebas atau tidak terdapat korelasi antar sesamanya, sehingga

dapat digunakan untuk mengatasi tidak terpenuhinya asumsi non-multikolinieritas pada logistik.

Regresi PLS-GLR didasarkan pada pembentukan komponen dengan prosedur iterasi menggunakan metode *Maximum Partial Likelihood Estimation* di mana metode ini merupakan metode pendugaan parameter yang digunakan dalam regresi.

Dalam PLS, matriks komponen $\mathbf{t}$ merupakan kombinasi linier dari peubah prediktor, dan ditulis dengan persamaan :

$$\mathbf{t} = \mathbf{XW} \quad (2.31)$$

di mana $\mathbf{W}_{(p \times m)}$ adalah matriks pembobot.

Matriks peubah prediktor $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ merupakan matriks yang menentukan peubah respon $\mathbf{Y}$ dan berfungsi untuk membentuk $\mathbf{t}_h$ komponen PLS yang ortogonal.

Dalam regresi PLS pembentukan $\mathbf{t}_h$ komponen untuk $h = 1, 2, \dots, n$ dapat dituliskan sebagai persamaan :

$$\mathbf{t}_h = w_{1h}\mathbf{x}_1 + w_{2h}\mathbf{x}_2 + \dots + w_{ph}\mathbf{x}_p \quad (2.32)$$

di mana w_{jh} merupakan normalisasi dari koefisien a_{jh}

$$w_{jh} = \frac{a_{jh}}{\|a_{jh}\|} \quad (2.33)$$

Untuk komponen $\mathbf{t}_1$, $\mathbf{x}_{j1} = \mathbf{x}_j$ dengan $\mathbf{x}_j$ adalah vektor peubah prediktor awal. Pembentukan komponen dilakukan dengan jalan mendapatkan koefisien a_{jh} yang signifikan pada hasil regresi dan pembentukan komponen berakhir jika sudah tidak ada lagi koefisien a_{jh} yang signifikan. Syarat pembentukan komponen PLS adalah adanya koefisien a_{jh} yang signifikan, jika diketahui tidak ada koefisien a_{jh} yang signifikan pada regresi maka pembentukan komponen PLS tidak dapat dilakukan.

Prosedur pembentukan komponen $\mathbf{t}_h$ untuk $h = 1, 2, \dots, m$ pada regresi dapat diuraikan sebagai berikut :

1. Pembentukan komponen $\mathbf{t}_1$ ($h = 1$)
 - a) Mendapatkan koefisien regrasi logistik a_{j1} dengan jalan meregresikan $\mathbf{x}_{j1}$ atau $\mathbf{x}_j$ (peubah prediktor awal) terhadap $\mathbf{Y}$ secara parsial,
 - b) Jika didapatkan koefisien a_{j1} yang signifikan maka dilakukan normalisasi koefisien a_{j1} dengan menggunakan persamaan (2.33),

sedangkan jika koefisien a_{j1} tidak signifikan maka ditetapkan bahwa nilai koefisien a_{j1} bernilai nol.

$$w_{j1} = \begin{cases} 0, & \text{untuk sig. } a_{j1} > 0.05 \\ \frac{a_{j1}}{\|a_{j1}\|}, & \text{untuk sig. } a_{j1} \leq 0.05 \end{cases}$$

- c) Menghitung skor komponen t_1 dari :

$$t_1 = Xw_i^*$$

$$t_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}(y, x_j)^2 x_j}} \quad (2.34)$$

$$t_1 = w_{11}x_1 + w_{21}x_2 + \dots + w_{p1}x_{p1} \quad (2.35)$$

2. Pembentukan komponen t_2 ($h = 2$)

- a) Membentuk vektor x_{j2} yang merupakan residual dari regresi linier antara masing-masing x_{j1} dengan t_1 .

Pada saat $h = 2$, maka $x_{j(h-1)} = x_{11}, x_{21}, \dots, x_{p1}$.

$$x_{1(h-1)} = x_{11} = p_{11}t_1 + x_{12}$$

$$x_{2(h-1)} = x_{21} = p_{21}t_1 + x_{22}$$

$$\vdots$$

$$x_{p(h-1)} = x_{p1} = p_{p1}t_1 + x_{p2}$$

- b) Menghitung koefisien a_{j2} dengan jalan melakukan regresi logistik x_{j2} dan t_1 terhadap Y secara parsial.
- c) Jika didapatkan koefisien a_{j2} signifikan, maka dilakukan normalisasi dengan persamaan (2.39), sedangkan jika didapatkan nilai koefisien a_{j2} tidak signifikan maka ditetapkan bahwa nilai koefisien a_{j2} bernilai nol.

$$w_{j2} = \begin{cases} 0, & \text{untuk sig. } a_{j2} > 0.05 \\ \frac{a_{j2}}{\|a_{j2}\|}, & \text{untuk sig. } a_{j2} \leq 0.05 \end{cases}$$

- d) Menghitung skor komponen t_2 dari :

$$t_2 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}(y, x_j)^2}} \text{cov}(y, x_j) x_{1j} \quad (2.36)$$

$$t_2 = w_{12}x_{12} + w_{22}x_{22} + \dots + w_{p2}x_{p2} \quad (2.37)$$

3. Pembentukan komponen $\mathbf{t}_h (h = 3, \dots, m)$

- a) Membentuk vektor $\mathbf{x}_{jh}$ yang merupakan residual dari regresi biasa antara masing-masing $\mathbf{x}_{j(h-1)}$ dengan $\mathbf{t}_{h-1}$. Pada saat $h = 3$, maka $\mathbf{x}_{j(h-1)} = \mathbf{x}_{12}, \mathbf{x}_{22}, \dots, \mathbf{x}_{p2}$. Pada saat $h = 4$, maka $\mathbf{x}_{j(h-1)} = \mathbf{x}_{13}, \mathbf{x}_{23}, \dots, \mathbf{x}_{p3}$, dan seterusnya sampai $h = m$, maka $\mathbf{x}_{j(h-1)} = \mathbf{x}_{1\ m-1}, \mathbf{x}_{2\ m-1}, \dots, \mathbf{x}_{p\ m-1}$.

$$\mathbf{x}_{1(h-1)} = p_{11} \mathbf{t}_1 + p_{12} \mathbf{t}_2 + \dots + p_{1(h-1)} \mathbf{t}_{(h-1)} + \mathbf{x}_{1h}$$

$$\mathbf{x}_{2(h-1)} = p_{21} \mathbf{t}_1 + p_{22} \mathbf{t}_2 + \dots + p_{2(h-1)} \mathbf{t}_{(h-1)} + \mathbf{x}_{2h}$$

$$\vdots$$

$$\mathbf{x}_{p(h-1)} = p_{p1} \mathbf{t}_1 + p_{p2} \mathbf{t}_2 + \dots + p_{p(h-1)} \mathbf{t}_{(h-1)} + \mathbf{x}_{ph}$$

- b) Menghitung nilai koefisien a_{jh} dengan jalan melakukan regresi logistik $\mathbf{Y}$ dengan $\mathbf{x}_{jh}$ dan $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{h-1}$ secara parsial. Jika didapatkan nilai koefisien a_{jh} signifikan maka dilakukan normalisasi, sedangkan jika nilai koefisien a_{jh} yang tidak signifikan ditetapkan bahwa nilai koefisien a_{jh} adalah nol.

$$w_{jh} = \begin{cases} 0, & \text{untuk } \text{sig. } a_{jh} > 0.05 \\ \frac{a_{jh}}{\|a_{jh}\|}, & \text{untuk } \text{sig. } a_{jh} \leq 0.05 \end{cases}$$

- c) Menghitung skor komponen $\mathbf{t}_h$ dari :

$$\mathbf{t}_h = w_{1h} \mathbf{x}_{1h} + w_{2h} \mathbf{x}_{2h} + \dots + w_{ph} \mathbf{x}_{ph} \quad (2.38)$$

Metode PLS-GLR (*Partial Least Square Generalized Linier Regression*) pada dasarnya bertujuan untuk mengetahui bentuk hubungan antara peubah prediktor (X) dengan peubah respon (Y). Berdasarkan hal tersebut perlu dilakukan transformasi dari hasil pembentukan komponen kedalam bentuk asal. Hasil transformasi ditunjukkan sebagai berikut :

2.8.1 Pendeteksian Multikolinieritas

Ghozali (2009) menerangkan ada beberapa cara untuk melihat multikolinieritas yaitu *tolerance* dan *Variance Inflation Factor* (VIF). Kedua ukuran ini menunjukkan setiap peubah independen manakah yang dijelaskan oleh peubah independen lainnya.

Menurut (Gujarati, 2010) VIF untuk koefisien regresi- j didefinisikan sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.41)$$

dengan R_j^2 adalah koefisien determinasi antara X_j dengan peubah prediktor lainnya pada persamaan atau model dugaan regresi, di mana $j = 1, 2, \dots, p$.

VIF menunjukkan bagaimana ragam dari sebuah penduga ditingkatkan oleh keberadaan multikolinieritas. Apabila nilai koefisien determinasi mendekati 1, maka nilai VIF akan bertambah besar. Hal ini akan terjadi jika suatu peubah prediktor mempunyai hubungan atau kolinieritas yang kuat dengan peubah prediktor lain.

Tolerance mengukur keragaman peubah prediktor terpilih yang tidak dijelaskan oleh peubah prediktor lainnya. Jadi *Tolerance* yang rendah sama dengan nilai VIF yang tinggi (karena $VIF = 1/\text{tolerance}$).

Suatu peubah prediktor dikatakan terjadi multikolinieritas jika $VIF > 5$, artinya dalam model tidak terjadi multikolinieritas jika nilai koefisien determinasi antar peubah prediktor hanya mencapai 0,8 maka jika nilai VIF lebih besar dari 10 maka terdapat masalah multikolinieritas yang serius pada data.

2.8.2 Pengaruh Multikolinieritas

Hosmer dan Lemeshow (2000) mengatakan bahwa seperti pada regresi logistik sensitif dengan adanya kolinieritas. Kolinieritas mengindikasikan bahwa terdapat hubungan ketergantungan antar peubah prediktor.

Pada kasus regresi linier, pendugaan parameter dengan metode kuadrat terkecil menjadi tidak tepat, dapat diuraikan sebagai berikut:

$$\hat{\beta} = [X'X]^{-1}X'Y \quad (2.42)$$

$X'X$ merupakan matriks ragam peragam, dari teori matriks jika paling sedikit terdapat satu baris/lajur dari X yang unsur-unsurnya merupakan kombinasi linier dari unsur-unsur baris/lajur lainnya, maka

$X'X$ merupakan matriks singular ($|X'X| = 0$). Matriks singular tidak mempunyai matriks kebalikan yang bersifat unik, sehingga $\hat{\beta}$ tidak dapat ditentukan secara pasti. Matriks $X'X$ merupakan matriks singular jika paling sedikit dua peubah X berkorelasi sempurna (misalnya korelasi antara X_1 dan X_2 , r_{12} sama dengan 1 atau -1). Begitu juga pada pendugaan parameter model regresi logistik, matriks ragam peragamnya adalah $[X'VX]^{-1}$ yang diasumsikan non singular, apabila di antara peubah X terdapat korelasi maka $|X'VX| = 0$

Multikolinieritas berpengaruh terhadap salah satu penduga koefisien regresi ($\hat{\beta}$), yaitu nilainya yang semakin besar bila terjadi multikolinieritas. Hal ini dapat dilihat dari matriks $[X'VX]^{-1}$ yang dibutuhkan dalam menentukan salah satu baku bagi pendugaan koefisien regresi logistik melalui persamaan:

$$Se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} \quad (2.43)$$

di mana:

$Var(\hat{\beta}_j)$ = elemen ke- j dari diagonal utama matriks $[X'VX]^{-1}$

Pada kasus multikolinieritas pada model logistik $X'VX$ bersifat singular atau mendekati singular ($|X'VX| \rightarrow 0$), sehingga hasil dari matriks $[X'VX]^{-1}$ yang diperoleh dari persamaan :

$$[X'VX]^{-1} = \frac{1}{|X'VX|} adj(X'VX) \quad (2.44)$$

akan semakin besar atau mendekati tak hingga. Hal ini menyebabkan nilai salah satu baku ($Se(\hat{\beta}_j)$) akan semakin besar juga. Adanya multikolinieritas dapat menghasilkan nilai penduga koefisien regresi, tetapi presisinya semakin rendah dengan semakin eratny hubungan antar peubah prediktor tersebut.

Semakin besar nilai $Se(\hat{\beta}_j)$ maka nilai statistik W pada uji wald semakin kecil, hal tersebut dapat dilihat dari rumus W hitung untuk menguji koefisien regresi logistik secara parsial untuk X_j sebagai berikut:

$$W_{hit} = \frac{|\hat{\beta}|}{Se(\hat{\beta}_j)} \quad (2.45)$$

$Se(\hat{\beta}_j)$ semakin besar jika X_j paling sedikit satu peubah X berkorelasi terhadap peubah lainnya, sehingga peluang $P[\chi^2 > W^2] < \alpha$ akan semakin menurun. Akibatnya peluang untuk menerima $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ semakin meningkat. Jadi adanya multikolinieritas akan menurunkan kekuatan uji dari uji Wald (Hosmer dan Lemeshow, 2000).

2.9 Kriteria Pemilihan Model Terbaik

2.9.1 Percent Correct Predictions (PCP).

Apabila terdapat beberapa model layak yang terbentuk dari beberapa metode, maka dibutuhkan suatu kriteria untuk memilih model yang terbaik di antara beberapa model yang terbentuk tersebut. Menurut Garson (2011) peluang kejadian dapat dihitung dengan membuat tabel klasifikasi *actual* dan *predicted group* yang dinamakan dengan *Percent Correct Predictions* (PCP) seperti pada tabel 2.1.

Tabel 2.1. Klasifikasi *actual* dan *predicted group*.

Anggota Aktual	Anggota Prediksi		Jumlah
	π_1	π_2	
π_1	n_{11}	$n_{12} = n_{1.} - n_{11}$	$n_{1.}$
π_2	$n_{21} = n_{2.} - n_{22}$	n_{22}	$n_{2.}$
Jumlah	$n_{.1}$	$n_{.2}$	N

Keterangan :

n_{11} : banyaknya pengamatan dari π_1 tepat diklasifikasi sebagai π_1

n_{12} : banyaknya pengamatan dari π_1 salah diklasifikasi sebagai π_2

n_{22} : banyaknya pengamatan dari π_2 tepat diklasifikasi sebagai π_2

n_{21} : banyaknya pengamatan dari π_2 salah diklasifikasi sebagai π_1

$$PCP = 1 - \left[\frac{n_{12} + n_{21}}{n_1 + n_2} \times 100\% \right] \quad (2.46)$$

Semakin besar nilai PCP yang dihasilkan maka semakin baik model yang diperoleh.

2.9.2 AIC (Akaike Information Criteria) dan BIC (Bayesian Information criteria)

Apabila terdapat beberapa model layak yang terbentuk dari beberapa metode, maka dibutuhkan suatu kriteria untuk memilih model yang terbaik di antara beberapa model yang terbentuk tersebut.

Chatterjee dan Hadi (2006) menyatakan bahwa AIC dan BIC dapat digunakan untuk mempertimbangkan model yang paling cocok di antara beberapa model logistik yang terbentuk.

Dalam model regresi logistik, rumus AIC adalah :

$$AIC = -2(\log\text{likelihood of the fitted model}) + 2p \quad (2.47)$$

di mana, p = banyaknya peubah prediktor di dalam model. Model dengan nilai AIC terkecil merupakan model yang paling baik.

Selain AIC ada juga ukuran kebaikan model lainnya yaitu BIC yang di definisikan oleh persamaan berikut :

$$BIC = -2(\log\text{likelihood of the fitted model}) + p \log n \quad (2.48)$$

di mana n = banyak contoh. Dalam BIC pemilihan model juga sama dengan AIC. BIC akan memberikan hasil yang baik ketika $n > 8$. Oleh sebab itu pemilihan model terbaik cenderung menggunakan nilai AIC.

2.10 Tinjauan Non Statistika

Usaha Mikro menurut Undang-Undang No 20 tahun 2008 tentang Usaha, Kecil dan Menengah (UKM) adalah usaha produktif milik orang perorangan dan/atau badan usaha perorangan yang memenuhi kriteria Usaha Mikro sebagaimana diatur dalam Undang-Undang tersebut.

Usaha mikro mempunyai peran yang penting dalam pembangunan ekonomi, karena intensitas tenaga kerja yang relatif lebih tinggi dan investasi yang lebih kecil, sehingga usaha mikro lebih fleksibel dalam menghadapi dan beradaptasi dengan perubahan pasar. Disamping itu tingkat penciptaan lapangan kerja lebih tinggi pada usaha mikro dari pada yang terjadi di perusahaan besar.

Pada dasarnya masing-masing bank mempunyai tata cara, persyaratan dan prosedur permohonan kreditnya sendiri-sendiri, namun tetap secara konsisten mengacu pada peraturan perundangan yang

berlaku bagi kalangan perbankan, terutama yang berkait dengan penerapan prinsip kehati-hatian. Secara garis besar prosedur untuk memperoleh kredit pada bank umum adalah sebagai berikut :

- a. Mengisi formulir aplikasi (permohonan kredit, data dan informasi perusahaan).
- b. Melengkapi persyaratan formulir permohonan kredit dengan dokumen-dokumen (data historis perusahaan, data proyeksi dan data jaminan).
- c. Analisis Kelayakan Kredit

Setiap permohonan kredit yang diajukan oleh calon debitur harus segera diproses melalui penilaian dan selanjutnya diberikan keputusannya oleh bank. Penilaian diwujudkan dalam bentuk pembuatan analisis kredit.

Penilaian atau analisis kredit mempunyai aspek yang besar pengaruhnya terhadap kelangsungan hidup bank. Tanpa penganalisaan yang cermat maka dapat mengakibatkan kerugian bagi pihak bank. Inilah yang mendasari Bank Mandiri secara khusus memiliki karakteristik penilaian dalam analisis kreditnya yang masih berhubungan dengan metode kualitatif maupun kuantitatif pada umumnya yang dilakukan oleh perbankan.

Dalam mengambil keputusan diterima atau ditolaknya permohonan pengajuan kredit, ada faktor-faktor yang mempengaruhi. Faktor tersebut terbagi atas faktor *social* dan faktor *financial* . Dari faktor sosial antara lain yaitu jumlah tanggungan keluarga yaitu banyaknya jumlah anggota keluarga yang menjadi tanggungan debitur. Sedangkan dari faktor finansial dapat dilihat dari lama usaha berdiri yaitu lama usaha berdiri yang telah dijalankan oleh debitur. Omzet usaha yaitu keseluruhan penjualan dari usaha yang dijalankan oleh debitur rata-rata per bulannya. Nilai agunan kredit yaitu objek yang dibiayai yang dijalankan oleh debitur rata – rata perbulannya Current Ratio (CR) yaitu perbandingan antara aktiva lancar dan hutang lancar. Debt service ratio (DSR) yaitu perbandingan antara besarnya angsuran dengan penghasilan, diukur dengan satuan jutaan rupiah. Biaya hidup yaitu perbandingan antara biaya hidup dibandingkan dengan laba maksimal 65% dari laba bersih perbulan dan dihitung dalam satuan rupiah. Dari semua peubah diduga menggunakan analisis regresi logistik untuk menentukan layak tidaknya dana tersebut diberikan kepada calon debitur.

BAB III BAHAN DAN METODE

3.1 Data Penelitian

Data yang digunakan adalah data sekunder yang di dapat dari MIS PT.Bank Mandiri (Persero) KCB Malang yang menangani Kredit Usaha Mikro (KUM). Pada data terlampir peubah peubah yang mempengaruhi keputusan pemberian kredit terhadap pengusaha mikro yaitu faktor sosial dan faktor finansial. Dalam pemodelan dengan metode PCLR_(s) dan PLS-GLR ingin deketahui berapa peluang keputusan di terima atau ditolaknya pengajuan aplikasi kredit oleh debitur. Adapun rinciannya adalah sebagai berikut:

Peubah respon:

- Y = 1 Jika pengajuan kredit diterima
- Y = 0 Jika pengajuan kredit ditolak

Peubah prediktor :

1. X1 : DSR yaitu *Dept Service Ratio*
2. X2 : Omzet Penjualan
3. X3 : CR yaitu *Current Ratio*
4. X4 : Biaya Hidup
5. X5 : Nilai Angunan Kredit
6. X6 : Lama usaha
7. X7 : Jumlah Tanggungan (anak dan istri)

3.2 Analisis Data

3.2.1 Metode Pembentukan Model Regresi Logistik dengan PCLR_(s)

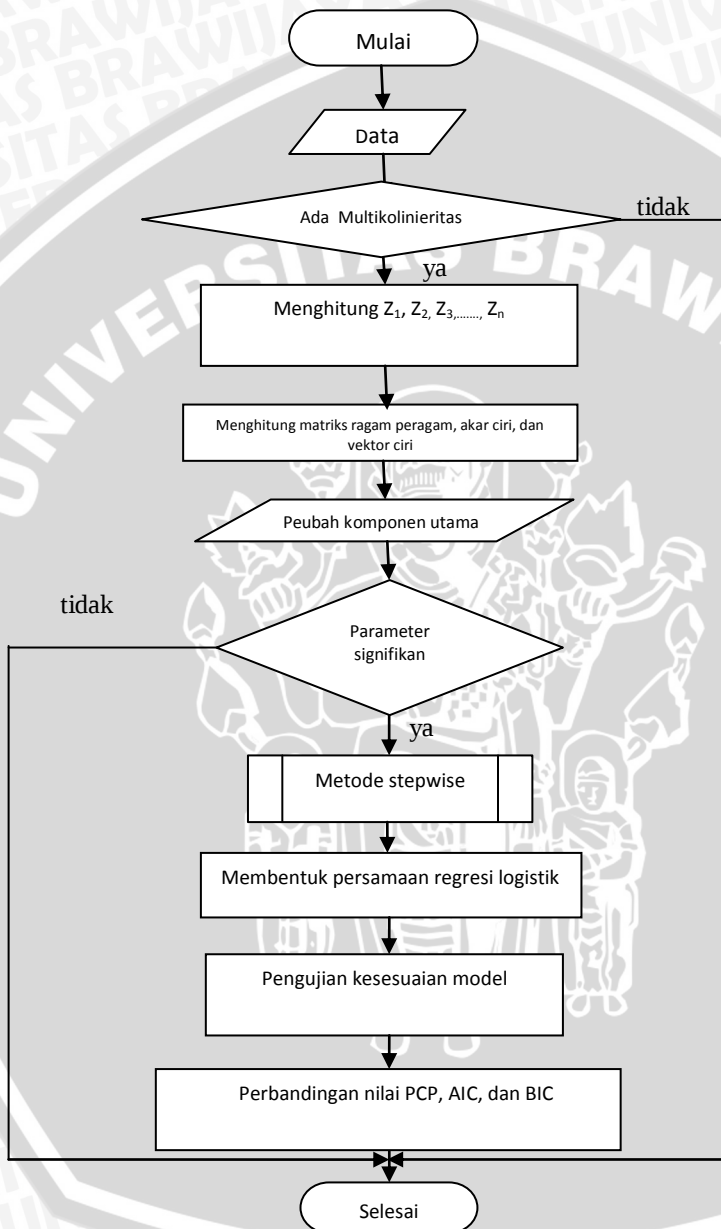
1. Pendeteksian multikolinieritas dengan persamaan (2.41)
2. Menghitung Menghitung $Z_1, Z_2, Z_3, \dots$ dengan cara membakukan $X_1, X_2, \dots, X_p$ (2.30)
3. Menghitung matriks ragam peragam S dengan persamaan (2.23)
4. Menghitung akar ciri (nilai eigen) dengan persamaan (2.24)
5. Menghitung vektor ciri (vektor eigen) dengan persamaan (2.28)
6. Menghitung skor peubah komponen utama dengan persamaan (2.29).
7. Menggunakan metode *stepwise* untuk memilih komponen utama yang masuk dalam model dan tidak saling berkorelasi. Langkah-langkah yang ditempuh terdapat pada Gambar 3.2.

8. Membentuk model persamaan regresi logistik dengan peubah respon Y dan peubah prediktor $W_{(1)}, W_{(2)}, \dots, W_{(s)}$ yang didapat dari metode *stepwise*.
9. Melakukan pengujian kesesuaian model dengan statistik uji *Pearson* dan *Deviance* sesuai dengan persamaan (2.43) dan (2.44).
10. Mendapatkan *Percent Correct Predictions* (PCP), AIC dan BIC dengan persamaan (2.46), (2.47) dan (2.48) kemudian dibandingkan dengan nilai yang didapatkan dari metode PLS-GLR

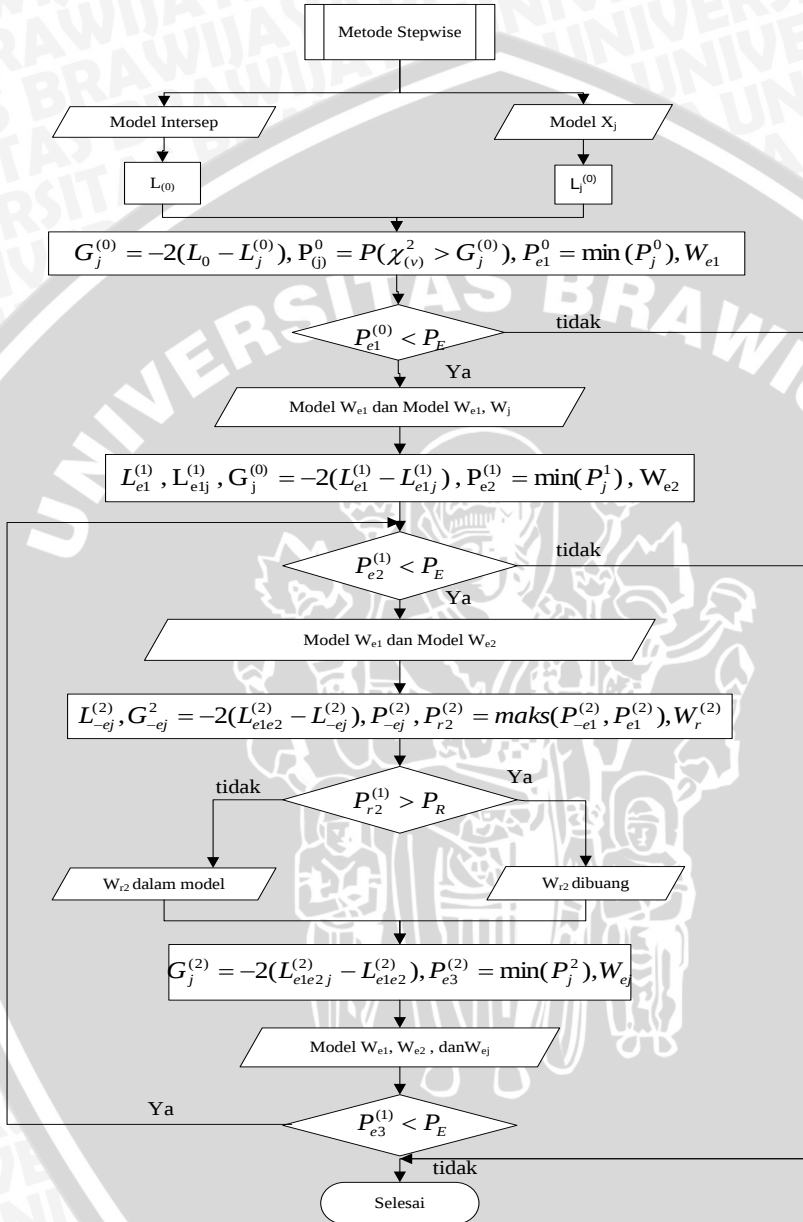
3.2.2 Metode Pembentukan Model Regresi Logistik dengan PLS-GLR

1. Pendeteksian multikolinieritas dengan persamaan (2.41).
2. Menghitung $X_1^*, X_2^*, \dots, X_p^*$ dengan cara membakukan $X_1, X_2, \dots, X_p$ (2.30)
3. Menghitung koefisien regresi logistik a_{ij}
4. Normalkan vektor kolom a_i yang berisi a_{ij} . Dengan persamaan (2.35).
5. Hitung komponen menghitung komponen PLS pertama yaitu t_1
6. Membentuk komponen t_h ($h = 1, 2, \dots, m$) dengan mendapatkan koefisien a_{jh} , jika didapatkan koefisien yang signifikan maka dilakukan normalisasi dengan persamaan (2.36)
7. Menghitung koefisien regresi logistik skor komponen t_h terhadap peubah respon Y
8. Membandingkan nilai PCP AIC dan BIC

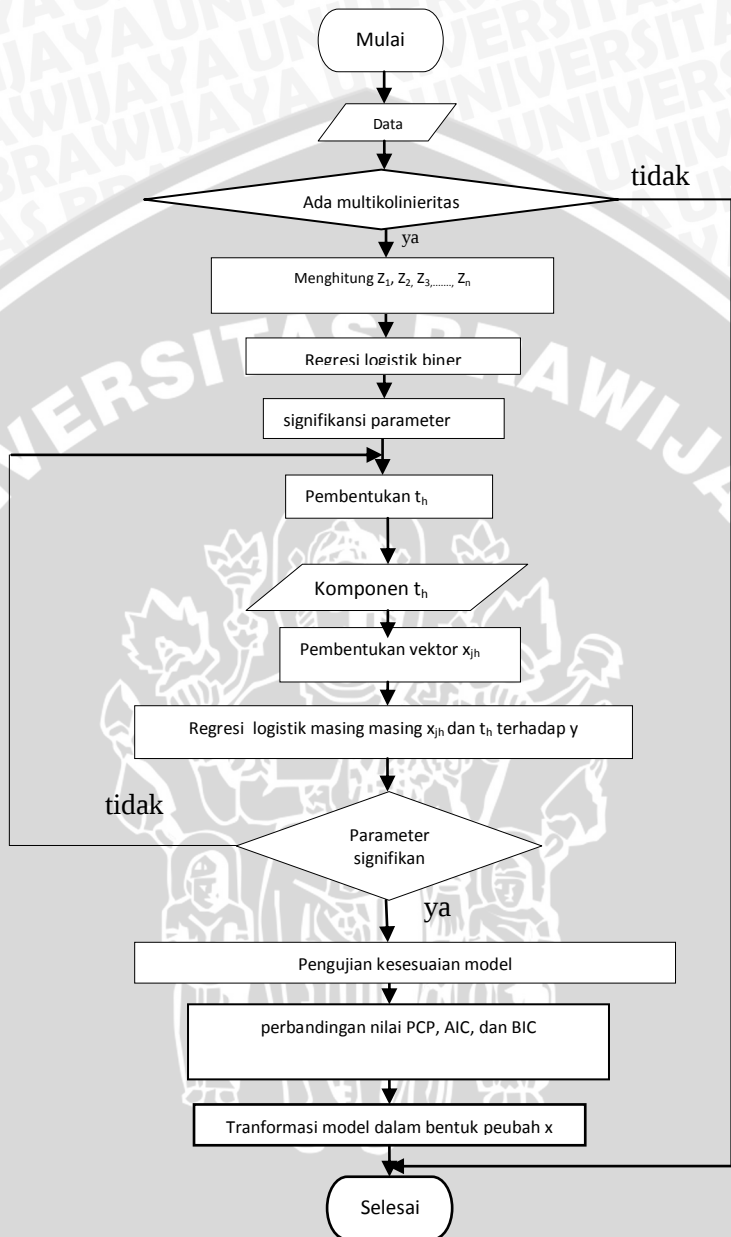
Tahapan dalam pemilihan model regresi logistik melalui metode PCLR_(s) dan PLS-GLR dikerjakan dengan bantuan *software* Minitab 14, SPSS 16.0 dan Microsoft *Excel*.



Gambar 3.1 Diagram alir pembentukan model regresi logistik dengan $PCLR_{(s)}$



Gambar 3.2 Diagram alir metode stepwise.



Gambar 3.3 Diagram alir pembentukan model regresi logistik dengan metode PLS-GLR

UNIVERSITAS BRAWIJAYA

