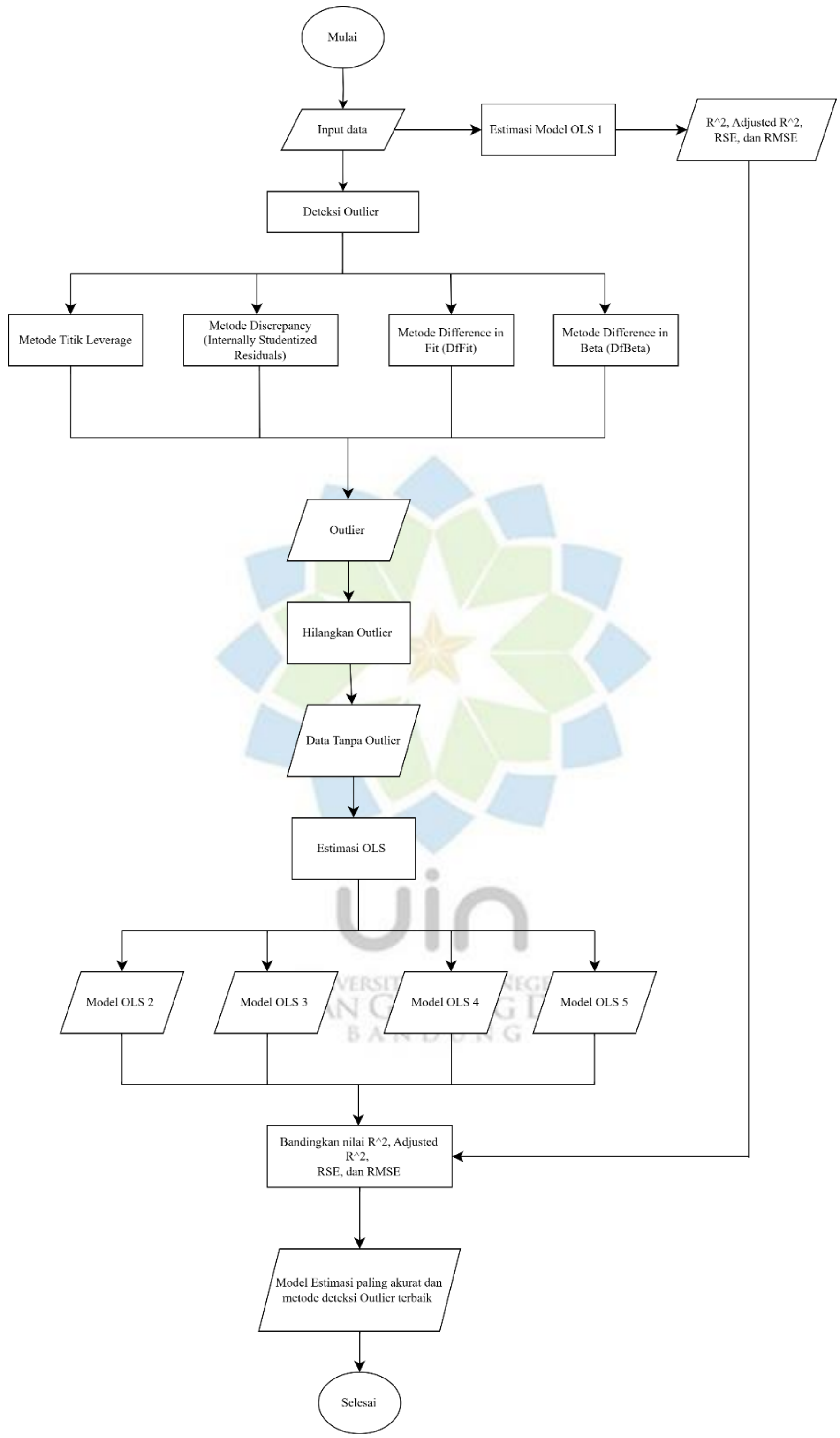
BAB III

PENDETEKSIAN OUTLIER PADA DATA LONGITUDINAL DENGAN METODE DFFIT DAN DFBETA

Pada data longitudinal terdapat penggabungan antara data *time series* dan data *cross section* masing-masing data dimodelkan dalam model *commont effect* yang didapatkan dari proses hasil estimasi OLS yang pertama untuk mendapatkan hasil residual minimum sehingga mendapatkan hasil estimasi yang akurat. Adapun, ketika nilai *error*-nya berada di angka yang sangat besar maka perlu dilakukannya proses pendeteksian *outlier* Proses pendeteksian *outlier* mempunyai tujuan untuk mengidentifikasi dan menangani data yang secara signifikan berbeda dari pola umum yang diamati pada individu atau waktu tertentu. Sehingga hasil yang diperoleh mencerminkan realitas yang akurat dan dapat diandalkan. Proses pendeteksian *outlier* yang dilakukan pada penelitian ini ada empat tahapan penting yaitu :

1. Menentukan nilai titik *leverage*
2. Menentukan nilai *Internally Studentized Residuals*
3. Menentukan nilai *DfFit*
4. Menentukan nilai *DfBeta* pada masing-masing koefisien

Setelah menentukan semua nilai pada keempat metode deteksi *outlier* tersebut dilanjutkan dengan proses mengidentifikasi *outlier* dengan kriteria nilai *cutoff* yang berbeda pada tiap metode sehingga menghasilkan data *outlier* yang berbeda juga pada masing-masing metode. Setelah didapatkan data yang termasuk *outlier* dilanjutkan dengan proses estimasi OLS setelah menghilangkan *outlier* tujuannya agar mendapatkan hasil koefisien determinasi dan *Adjusted* koefisien determinasi paling tinggi dan juga menghasilkan nilai RSE dan RMSE paling minimum sebagai kriteria untuk menentukan metode deteksi *outlier* terbaik dan estimasi OLS paling akurat. Berikut merupakan langkah-langkah dari metode yang digunakan secara keseluruhan :



Gambar 3.1 *Flowchart* langkah-langkah deteksi *outlier* dan estimasi parameter OLS pada masing-masing metode

3.1 Struktur Model Data Longitudinal

Semua jenis penilaian berulang pada individu yang sama termasuk dalam analisis data longitudinal. Berikut merupakan struktur data longitudinal dengan nilai $i = 1,2,\dots,n$ banyaknya objek observasi dan nilai $t = 1,2,\dots,T$ banyaknya waktu/periode observasi yang dilakukan, serta $j = 1,2,\dots,K$ banyaknya koefisien dalam pengamatan yang dilakukan.

Tabel 3.1 Struktur Data Longitudinal

$\boldsymbol{Y_{i,t}}$	$\boldsymbol{t}$	$\boldsymbol{X_{1,i,t}}$	$\boldsymbol{X_{2,i,t}}$		$\boldsymbol{X_{j,i,t}}$
$Y_{1,1}$	1	$X_{1,1,1}$	$X_{2,1,1}$	…	$X_{K,1,1}$
$Y_{2,1}$	1	$X_{1,2,1}$	$X_{2,2,1}$	…	$X_{K,2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Y_{N,1}$	1	$X_{1,N,1}$	$X_{2,N,1}$	…	$X_{K,N,1}$
$Y_{1,2}$	2	$X_{1,1,2}$	$X_{2,1,2}$	…	$X_{K,1,2}$
$Y_{2,2}$	2	$X_{1,2,2}$	$X_{2,2,2}$	…	$X_{K,2,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Y_{N,2}$	2	$X_{1,N,1}$	$X_{2,N,2}$	…	$X_{K,N,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Y_{1,T}$	T	$X_{1,1,T}$	$X_{1,1,T}$	…	$X_{K,1,T}$
$Y_{2,T}$	T	$X_{1,2,T}$	$X_{2,2,T}$	…	$X_{K,2,T}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Y_{N,T}$	T	$X_{1,N,T}$	$X_{2,N,T}$	…	$X_{K,N,T}$

3.2 Model Data Longitudinal dan Estimasi Parameter OLS

Ada tiga kategori umum model data longitudinal yaitu *Commont Effect Model, Random Effect Model,* dan *Fixed Effect Model* akan tetapi model yang digunakan dalam penelitian ini hanya model *Commont Effect* yaitu model yang merupakan pendugaan seluruh data *time series* dan *cross section* digabungkan (*pooled*) untuk menduga parameternya dengan metode OLS (*Ordinary Least Square*) yang akan dijelaskan langkah-langkahnya sebagai berikut :

A. Model *Commont Effect* pada data longitudinal

Langkah-langkah menentukan model *commont effect* pada data longitudinal :

1. Tentukan matriks X_{jit} dengan i merupakan jumlah pengamatan (*cross section*) sebanyak $i = 1,2,\dots,N,$ t merupakan rentang waktu yang digunakan (*time series*) $t = 1,2,\dots,T$ dan untuk j merupakan banyaknya variabel prediktor yang digunakan berukuran $j = 1,\dots,K$ dengan ukuran $n \times p$

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix} \tag{3.1}$$

2. Tentukan nilai rata-rata dari variabel X dengan n adalah banyaknya data

$$\bar{X}_{jit} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{jit}}{n} \tag{3.2}$$

3. Tentukan matriks nilai variabel respon unit *cross section* ke-i untuk periode waktu ke-t berukuran $n \times 1$

$$Y_{it} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{N2} \end{bmatrix}, \dots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix} \tag{3.3}$$

4. Tentukan nilai rata-rata dari variabel Y

$$\bar{Y}_{it} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}}{n} \tag{3.4}$$

5. Tentukan nilai β_1 yang menunjukkan nilai koefisien dari variabel yang pertama

$$\beta_1 = \frac{\sum_{j=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{1it} - \bar{X}_{1it})\,(Y_{1t} - \bar{Y}_{1t})}{\sum_{j=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{1it} - \bar{X}_{1it})^2} \tag{3.5}$$

6. Tentukan nilai β_0 yang menunjukkan *intercept* atau konstanta unit *cross section*

$$\beta_0 = \bar{Y}_{it} - \beta_1 \bar{X}_{jit} \tag{3.6}$$

7. Tentukan matriks β_j yang merupakan vektor *slope* atau koefisien regresi sebanyak *K* variabel prediktor berukuran $p\ \times 1$

$$\beta_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \tag{3.7}$$

8. Tentukan matriks ε_{it} yang merupakan galat regresi unit *cross section* pada pengamatan ke-*i* untuk periode waktu ke-*t* berukuran $n \times 1$

$$\varepsilon_{it} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{N1} \end{bmatrix}, \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{N2} \end{bmatrix}, \dots, \begin{bmatrix} \varepsilon_{1\mathrm{T}} \\ \varepsilon_{2\mathrm{T}} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix} \tag{3.8}$$

9. Tetapkan persamaan *commont effect model* untuk $i = 1, i = 2, \dots., i = N$, untuk persamaan matriks pada observasi kesatu atau untuk $i = 1$ adalah

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T} \end{bmatrix} = \beta_0 + \begin{bmatrix} X_{111} & X_{211} & \cdots & X_{K11} \\ X_{112} & X_{212} & \cdots & X_{K12} \\ \vdots & \vdots & \ddots & \vdots \\ X_{11T} & X_{21T} & \cdots & X_{K1T} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T} \end{bmatrix} \tag{3.9}$$

persamaan matriks pada observasi kedua atau untuk $i = 2$ adalah

$$\begin{bmatrix} Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2T} \end{bmatrix} = \beta_0 + \begin{bmatrix} X_{121} & X_{221} & \cdots & X_{K21} \\ X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \ddots & \vdots \\ X_{12T} & X_{22T} & \cdots & X_{K2T} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2T} \end{bmatrix}$$

Lalu untuk persamaan matriks pada observasi ke-n atau untuk $i = N$ adalah

$$\begin{bmatrix} Y_{N1} \\ Y_{N2} \\ \vdots \\ Y_{NT} \end{bmatrix} = \beta_0 + \begin{bmatrix} X_{1\mathrm{N}1} & X_{2\mathrm{N}1} & \cdots & X_{KN1} \\ X_{1\mathrm{N}2} & X_{2\mathrm{N}2} & \cdots & X_{KN2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_{\mathrm{N}1} \\ \varepsilon_{\mathrm{N}2} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

10. Bentuk persamaan keseluruhan dari matriks diatas

$$Y_{it} = \beta_0 + \sum_{j=1}^{K} \beta_j X_{jit} + \varepsilon_{it} \tag{3.10}$$

Apabila dibuat dalam bentuk matriks maka

$$\begin{bmatrix} Y_{11} \\ Y_{22} \\ \vdots \\ Y_{NT} \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0N} \end{bmatrix} + \begin{bmatrix} X_{111} & X_{211} & \cdots & X_{K11} \\ X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

B. Estimasi Parameter dengan Metode *Ordinary Least Square*

Langkah-langkah estimasinya dijelaskan sebagai berikut :

1. Tetapkan matriks ε_{it} dan transpose matriks $\varepsilon_{it}{}^T$

$$\varepsilon_{it} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{N1} \end{bmatrix}, \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{N2} \end{bmatrix}, \dots, \begin{bmatrix} \varepsilon_{1\mathrm{T}} \\ \varepsilon_{2\mathrm{T}} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

untuk transpose matriknya adalah

$$\varepsilon_{it}{}^T = [\varepsilon_{11} \quad \varepsilon_{21} \quad \cdots \quad \varepsilon_{N1}], \dots, [\varepsilon_{1\mathrm{T}} \quad \varepsilon_{2\mathrm{T}} \quad \cdots \quad \varepsilon_{NT}]$$

2. Tetapkan nilai matriks Y_{it}

$$Y_{it} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{N2} \end{bmatrix}, \dots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

3. Tetapkan matriks X_{jit}

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

4. Tetapkan matriks β_j

$$\beta_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

5. Jika S merupakan total kuadrat *error* maka

$$\begin{aligned} S = \varepsilon_{it}{}^T \varepsilon_{it} &= [\varepsilon_{11} \quad \varepsilon_{22} \quad \cdots \quad \varepsilon_{NT}] \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix} \\ &= \varepsilon_{11}\varepsilon_{11} + \varepsilon_{22}\varepsilon_{22} + \cdots + \varepsilon_{NT}\varepsilon_{NT} = \sum \varepsilon_{it}{}^2 \\ &= (Y_{it} - X_{jit}\beta_j)^T (Y_{it} - X_{jit}\beta_j) \end{aligned} \tag{3.11}$$

6. Karena persamaan diatas merupakan bentuk skalar sehingga untuk nilai S dapat didefinisikan sebagai

$$\begin{aligned} S &= \left(Y_{it} - X_{jit}\right)^{T}\left(Y_{it} - X_{jit}\beta_j\right) \\ &= \left(Y_{it} - (X_{jit}\beta_j)^{T}\right)\left(Y_{it} - X_{jit}\beta_j\right) \\ &= ({Y_{it}}^{T} - {\beta_j}^{T}{X_{jit}}^{T})\left(Y_{it} - X_{jit}\beta_j\right) \\ &= {Y_{it}}^{T}Y_{it} - {Y_{it}}^{T}X_{jit}\beta_j - {X_{jit}}^{T}Y_{it}{\beta_j}^{T} + {\beta_j}^{T}{Y_{it}}^{T}X_{jit}\beta_j \\ &= {Y_{it}}^{T}Y_{it} - 2{\beta_j}^{T}{X_{jit}}^{T}Y_{it} + {\beta_j}^{T}{X_{jit}}^{T}X_{jit}\beta_j \end{aligned} \quad (3.11.1)$$

7. Minimumkan nilai S dengan dicari turunan pertama dari S terhadap β yaitu $\left.\frac{\vartheta(S)}{\vartheta\beta}\right|_{\beta=\widehat{\beta}}$ sehingga diperoleh persamaan normalnya yaitu :

$$-2{X_{jit}}^{T}Y_{it} + 2{X_{jit}}^{T}X_{jit}\hat{\beta}_j = 0$$

$$2{X_{jit}}^{T}X_{jit}\hat{\beta}_j = 2{X_{jit}}^{T}Y_{it}$$

$${X_{jit}}^{T}X_{jit}\hat{\beta}_j = {X_{jit}}^{T}Y_{it}$$

atau bisa juga ditulis :

$$\hat{\beta}_{OLS} = ({X_{jit}}^{T}X_{jit})^{-1}{X_{jit}}^{T}Y_{it} \quad (3.12)$$

C. Estimasi OLS setelah menghapus *outlier*

Langkah-langkah estimasinya dijelaskan sebagai berikut :

1. Tetapkan matriks ε_{it} dan transpose matriks-nya

$$\varepsilon_{it} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{(i-1)1} \\ \varepsilon_{(i+1)1} \\ \vdots \\ \varepsilon_{N1} \end{bmatrix}, \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{(i-1)2} \\ \varepsilon_{(i-1)2} \\ \vdots \\ \varepsilon_{N2} \end{bmatrix}, \dots, \begin{bmatrix} \varepsilon_{1T} \\ \varepsilon_{2T} \\ \vdots \\ \varepsilon_{(i-1)T} \\ \varepsilon_{(i-1)T} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

transpose matriknya adalah

$${\varepsilon_{it}}^{T} = [\varepsilon_{11} \quad \varepsilon_{21} \quad \cdots \quad \varepsilon_{(i-1)1} \quad \varepsilon_{(i+1)1} \quad \cdots \quad \varepsilon_{N1}], \dots, [\varepsilon_{1T} \quad \varepsilon_{2T} \quad \cdots \quad \varepsilon_{(i-1)T} \quad \varepsilon_{(i+1)T} \quad \cdots \quad \varepsilon_{NT}]$$

2. Tetapkan nilai matriks $Y_{(i)t}$

$$Y_{(i)t} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{(i-1)1} \\ Y_{(i+1)1} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{(i-1)2} \\ Y_{(i+1)2} \\ \vdots \\ Y_{N2} \end{bmatrix}, \dots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{(i-1)T} \\ Y_{(i+1)T} \\ \vdots \\ Y_{NT} \end{bmatrix} \qquad (3.13)$$

3. Tetapkan matriks $X_{j(i)t}$

$$X_{j(i)t} = \begin{bmatrix} X_{111} & X_{211} & & X_{K11} \\ X_{121} & X_{221} & & X_{K21} \\ \vdots & \vdots & & \vdots \\ X_{1(i-1)1} & X_{2(i-1)1} & \dots & X_{K(i-1)1} \\ X_{1(i+1)1} & X_{2(i+1)1} & & X_{K(i+1)1} \\ \vdots & \vdots & & \vdots \\ X_{1N1} & X_{2N1} & & X_{KN1} \end{bmatrix} \dots \begin{bmatrix} X_{11T} & X_{21T} & & X_{K1T} \\ X_{12T} & X_{22T} & & X_{K2T} \\ \vdots & \vdots & & \vdots \\ X_{1(i-1)T} & X_{2(i-1)T} & \dots & X_{K(i-1)T} \\ X_{1(i+1)T} & X_{2(i+1)T} & & X_{K(i+1)T} \\ \vdots & \vdots & & \vdots \\ X_{1NT} & X_{2NT} & & X_{KNT} \end{bmatrix} \qquad (3.14)$$

4. Tetapkan matriks β_j

$$\beta_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

5. Jika S merupakan total kuadrat *error* maka

$$\begin{aligned} S &= \varepsilon_{(i)t}{}^T \varepsilon_{(i)t} \\ &= \varepsilon_{11}\varepsilon_{11} + \varepsilon_{22}\varepsilon_{22} + \dots + \varepsilon_{NT}\varepsilon_{NT} \\ &= \sum \varepsilon_{(i)t}{}^2 \\ &= \left(Y_{(i)t} - X_{j(i)t}\beta_j\right)^T \left(Y_{(i)t} - X_{j(i)t}\beta_j\right) \end{aligned} \qquad (3.15)$$

6. Karena persamaan diatas merupakan bentuk skalar sehingga untuk nilai S dapat didefinisikan sebagai

$$\begin{aligned} S &= \left(Y_{(i)t} - X_{j(i)t}\right)^T \left(Y_{(i)t} - X_{j(i)t}\beta_j\right) \\ &= \left(Y_{(i)t} - (X_{j(i)t}\beta_j)^T\right)\left(Y_{(i)t} - X_{j(i)t}\beta_j\right) \\ &= \left(Y_{(i)t}{}^T - \beta_j{}^T X_{j(i)t}{}^T\right)\left(Y_{(i)t} - X_{j(i)t}\beta_j\right) \\ &= Y_{(i)t}{}^T Y_{(i)t} - Y_{(i)t}{}^T X_{j(i)t}\beta_j - X_{j(i)t}{}^T Y_{(i)t}\beta_j{}^T + \beta_j{}^T Y_{(i)t}{}^T X_{j(i)t}\beta_j \\ &= Y_{(i)t}{}^T Y_{(i)t} - 2\beta_j{}^T X_{j(i)t}{}^T Y_{(i)t} + \beta_j{}^T X_{j(i)t}{}^T X_{j(i)t}\beta_j \end{aligned} \qquad (3.15.1)$$

7. Minimumkan nilai S dengan dicari turunan pertama dari S terhadap β yaitu $\left.\frac{\vartheta(S)}{\vartheta\beta}\right|_{\beta=\hat{\beta}}$ sehingga diperoleh persamaan normalnya yaitu :

$$-2X_{j(i)t}{}^{T}Y_{(i)t} + 2X_{j(i)t}{}^{T}X_{j(i)t}\hat{\beta}_j = 0$$

$$2X_{j(i)t}{}^{T}X_{j(i)t}\hat{\beta}_j = 2X_{j(i)t}{}^{T}Y_{(i)t}$$

$$X_{j(i)t}{}^{T}X_{j(i)t}\hat{\beta}_j = X_{j(i)t}{}^{T}Y_{(i)t}$$

atau bisa didefinisikan dengan

$$\hat{\beta}_{OLS} = (X_{j(i)t}{}^{T}X_{j(i)t})^{-1}X_{j(i)t}{}^{T}Y_{(i)t} \tag{3.16}$$

3.3 Pendeteksian *Outlier*

Deteksi *outlier* dengan metode titik *leverage*, *internally studentized residuals*, *DfFit*, dan *DfBeta* bertujuan untuk mengidentifikasi pengamatan yang memiliki pengaruh terhadap model regresi, yang dapat mengganggu validitas hasil analisis. Titik *leverage* membantu mengungkap pengamatan yang secara signifikan berbeda dalam hal nilai prediktor, sedangkan *internally studentized residuals* menilai kesalahan prediksi relatif terhadap distribusi residual. *DfFit* mengukur pengaruh pengamatan terhadap nilai yang diprediksi, dan *DfBeta* menilai dampak pengamatan terhadap estimasi koefisien regresi. Penggabungan metode-metode ini memungkinkan diperoleh hasil yang lebih akurat dalam mendeteksi *outlier* yang memungkinkan merusak model hasil estimasinya.

3.3.1 Menghitung Titik *Leverage*

Langkah-langkah dalam menentukan nilai titik *leverage* didefinisikan sebagai berikut :

1. Tetapkan matriks X_{jit} dengan $j = 1, \dots, K, i = 1, \dots, N,$ dan $t = 1, \dots, T$

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

2. Tetapkan matriks tranformasi dari X_{jit}

$$X_{jit}{}^{T} = \begin{bmatrix} 1 & 1 & & 1 \\ X_{111} & X_{122} & & X_{1NT} \\ X_{211} & X_{222} & \dots & X_{2NT} \\ \vdots & \vdots & & \vdots \\ X_{K11} & X_{K22} & & X_{KNT} \end{bmatrix}$$

3. Tetapkan nilai $\left(X_{jit}{}^{T} X_{jit}\right)^{-1}$

$$\left(X_{jit}{}^{T} X_{jit}\right)^{-1} = \left(\begin{bmatrix} 1 & 1 & & 1 \\ X_{111} & X_{122} & & X_{1NT} \\ X_{211} & X_{222} & \dots & X_{2NT} \\ \vdots & \vdots & & \vdots \\ X_{K11} & X_{K22} & & X_{KNT} \end{bmatrix} \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix} \right)^{-1}$$

4. Tentukan matriks H

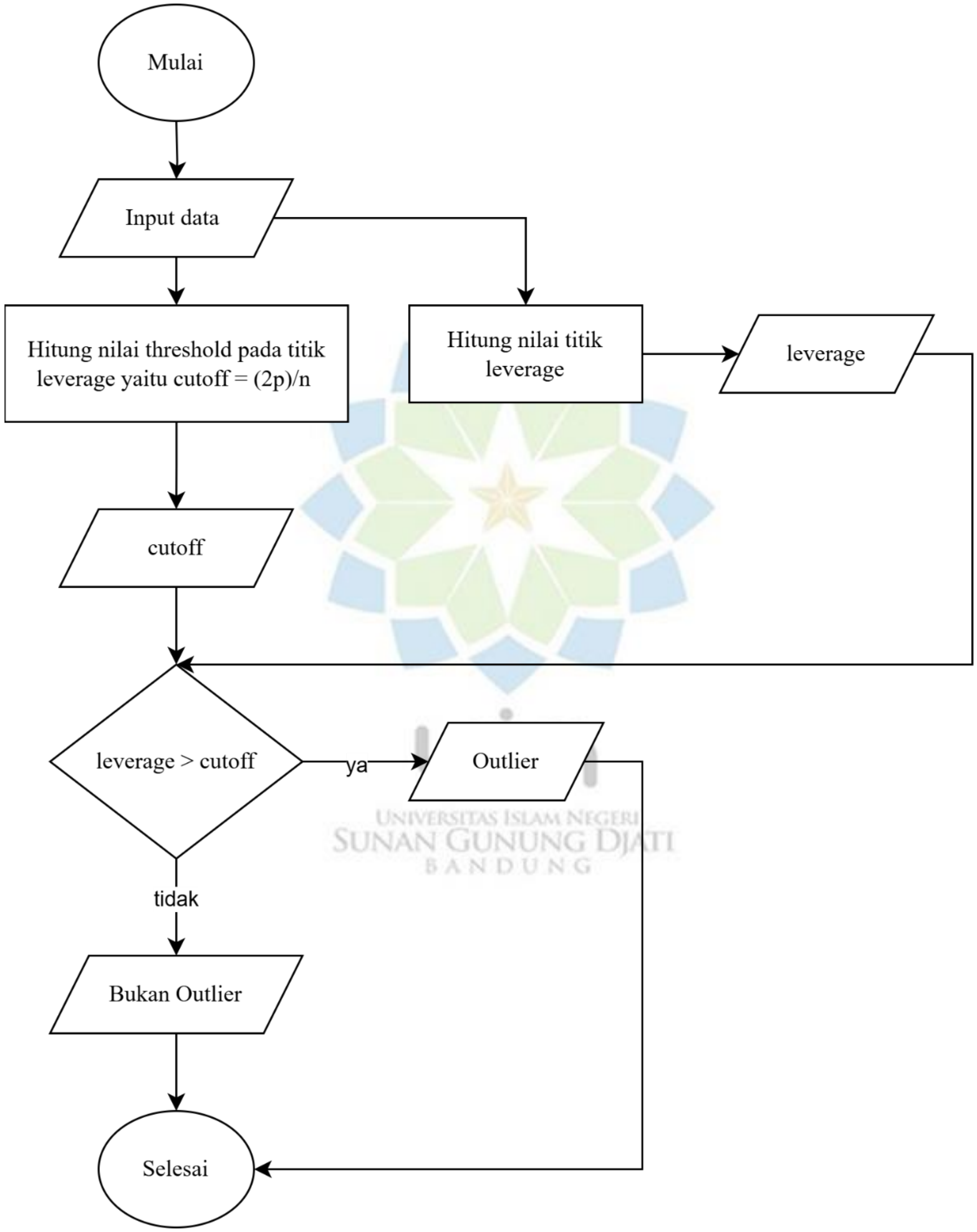
$$H = X_{jit}\left(X_{jit}{}^{T} X_{jit}\right)^{-1} X_{jit}{}^{T} \tag{3.15}$$

5. Nilai titik *leverage* untuk suatu observasi ke-i didefinisikan sebagai ini merupakan hasil esktrak dari elemen diagonal h_{ii}

$$h_{ii} = [H]_{ii} = x_{jit}{}^{T}\left(X_{jit}{}^{T} X_{jit}\right)^{-1} x_{jit} \tag{3.16}$$

6. Identifikasi *outlier* dengan metode nilai titik *leverage* yaitu menghitung kriteria $h_{ii} > \frac{2p}{n}$ dengan p adalah banyaknya parameter dan n adalah banyaknya objek pengamatan, maka data dikatakan sebagai *outlier.*

Langkah-langkah pencarian nilai titik *leverage* dapat dilihat pada diagram alir yang digambarkan pada Gambar 3.2



Gambar 3.2 *Flowchart* langkah-langkah deteksi *outlier* dengan metode titik *leverage*

3.3.2 Menghitung Nilai *Internally Studentized Residuals*

Internally studentized residuals merupakan jumlah dari banyaknya nilai kasus ke-i dengan standar deviasi dari residual kasus ke-i. Mengidentifikasi *outlier* dengan metode *internally studentized residuals* dijelaskan langkah-langkahnya sebagai berikut:

1. Tentukan matriks X pada pengamatan koefisien ke-j, observasi ke-i, dan waktu ke-t

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

2. Tetapkan matriks transpose dari X_{jit}

$$X_{jit}{}^{T} = \begin{bmatrix} 1 & 1 & & 1 \\ X_{111} & X_{122} & & X_{1NT} \\ X_{211} & X_{222} & \cdots & X_{2NT} \\ \vdots & \vdots & & \vdots \\ X_{K11} & X_{K22} & & X_{KNT} \end{bmatrix}$$

3. Tentukan matriks Y_{it}

$$Y_{it} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{N2} \end{bmatrix}, \dots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

4. Hitung $\bar{Y}$ dan $\bar{X}$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_{it}}{n} \text{ dan } \bar{X} = \frac{\sum_{i=1}^{n} X_{it}}{n}$$

5. Hitung $\hat{\beta}_1$

$$\hat{\beta}_1 = \left(X_{i1}{}^{T} X_{i1}\right)^{-1} X_{i1}{}^{T} Y_{it}$$

6. Tentukan $\hat{\beta}_0$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

7. Tentukan matriks $\hat{\beta}_j$

$$\widehat{\beta_J} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

8. Tentukan nilai $\hat{Y}_{it}$

$$\hat{Y}_{it} = \hat{\beta}_0 + \sum_{j=1}^{K} \hat{\beta}_j X_{ij} \tag{3.17}$$

9. Tentukan nilai titik *leverage* menggunakan rumus berikut :

$$h_{ii} = x_{jit}{}^T \left(X_{jit}{}^T X_{jit}\right)^{-1} x_{jit}$$

10. Tentukan nilai residual mengggunakan rumus berikut :

$$e_{it} = Y_{it} - \hat{Y}_{it} \tag{3.18}$$

11. Tentukan nilai residual kuadrat dari observasi ke-i

$$e_{it}^2 = \left(Y_{it} - \hat{Y}_{it}\right)^2$$

12. Tentukan nilai *residual sum square* (RSS)

$$RSS = \sum_{i=1}^{N} e_{it}^2 \tag{3.19}$$

13. Tentukan nilai estimasi standar deviasi dari residuals

$$\hat{\sigma}_{it} = \sqrt{\frac{RSS}{n-p}} \tag{3.20}$$

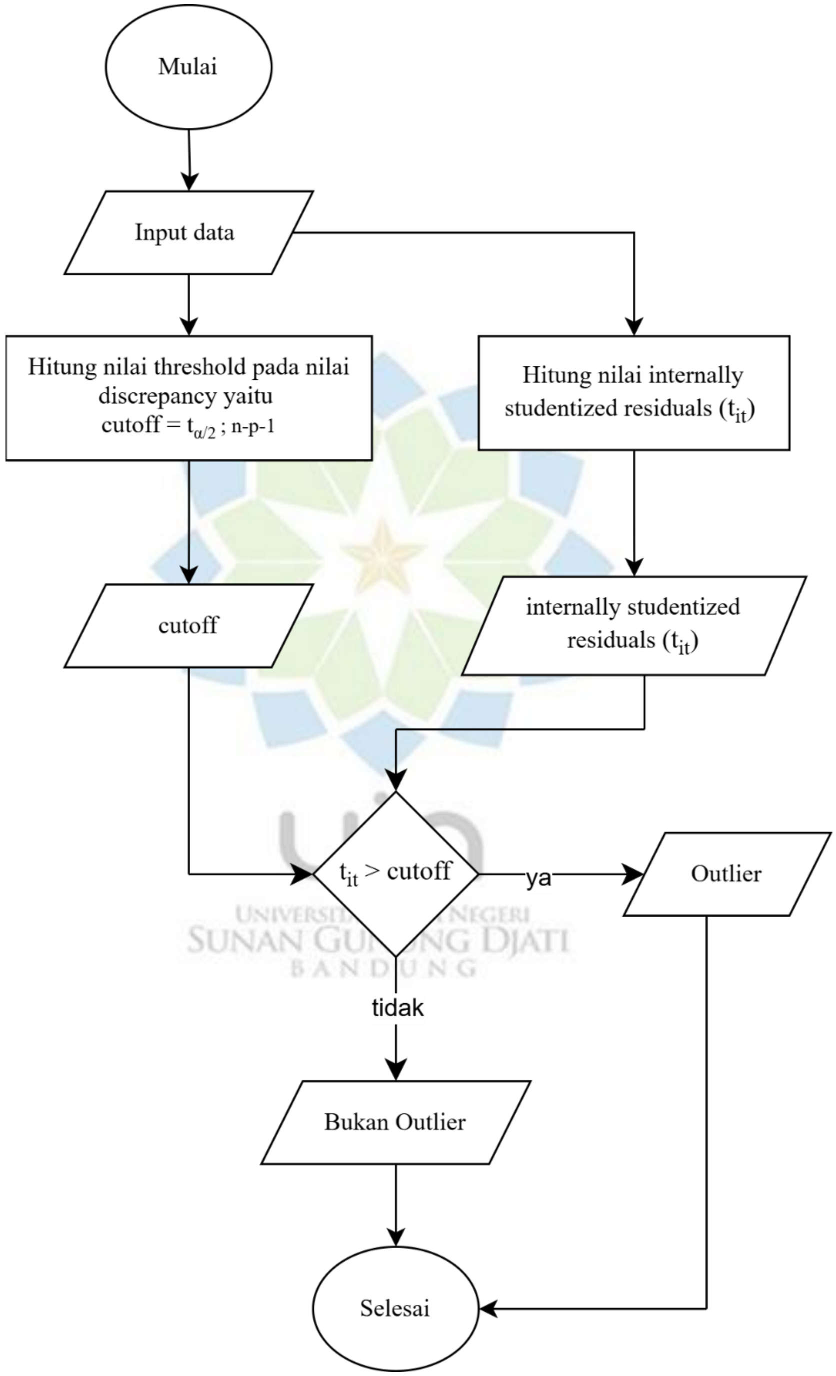
14. Tentukan nilai *internally studentized residual* untuk menentukan adanya *outlier*

$$t_{it} = \frac{e_{it}}{\hat{\sigma}_{it}\sqrt{1-h_{ii}}} \tag{3.21}$$

15. Setelah mendapatkan nilai dari *internally studentized residual* maka tentukan taraf signifikansinya dengan $\alpha = 0.05$ maka $|t_{it}| > t_{\frac{\alpha}{2};n-p-1}$

16. Nilai $t_{\frac{\alpha}{2};n-p-1}$ berada di kisaran $+2$ sampai -2 apabila hasil dari nilai *internally studentized residuals* melebihi dari nilai tersebut maka data dikatakan sebagai *outlier.*

Langkah-langkah pencarian nilai *internally studentized residuals* dapat dilihat pada diagram alir yang digambarkan pada Gambar 3.3



Gambar 3.3 *Flowchart* langkah-langkah deteksi *outlier* dengan metode *Internally Studentized Residuals*

3.3.3 Metode *Difference in Fit* (*DfFit*)

Mengidentifikasi *outlier* dengan metode *DfFit* berdasarkan penggunaan nilai *internally studentized residuals* dijelaskan langkah-langkahnya sebagai berikut :

1. Tentukan matriks X pada koefisien ke-j, observasi ke-i, dan waktu ke-t

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

2. Tetapkan matriks transpose dari X_{jit}

$$X_{jit}{}^{T} = \begin{bmatrix} 1 & 1 & & 1 \\ X_{111} & X_{122} & & X_{1NT} \\ X_{211} & X_{222} & \dots & X_{2NT} \\ \vdots & \vdots & & \vdots \\ X_{K11} & X_{K22} & & X_{KNT} \end{bmatrix}$$

3. Hitung $\bar{Y}$ dan $\bar{X}$

$$\bar{Y} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}}{n} \text{ dan } \bar{X} = \frac{\sum_{j=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{jit}}{n}$$

4. Tentukan matriks Y_{it}

$$Y_{it} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{N2} \end{bmatrix}, \dots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

5. Hitung $\hat{\beta}_1$

$$\hat{\beta}_1 = \left(X_{i1t}{}^{T} X_{i1t}\right)^{-1} X_{i1t}{}^{T} Y_{it}$$

6. Tentukan $\hat{\beta}_0$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

7. Tentukan matriks $\hat{\beta}_j$

$$\widehat{\beta_J} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

8. Tentukan nilai $\hat{Y}_{it}$

$$\hat{Y}_{it} = \hat{\beta}_0 + \sum_{j=1}^{K} \hat{\beta}_j X_{jit}$$

9. Tentukan nilai titik *leverage* menggunakan rumus berikut :

$$h_{ii} = x_i^T (X'X)^{-1} x_i$$

10. Tentukan nilai residual mengggunakan rumus berikut :

$$e_{it} = Y_{it} - \hat{Y}_{it}$$

11. Tentukan nilai residual kuadrat dari observasi ke-i

$$e_{it}^2 = \left(Y_{it} - \hat{Y}_{it}\right)^2$$

12. Tentukan nilai *residual sum square* (RSS)

$$RSS = \sum_{i=1}^{N} e_{it}^2$$

13. Tentukan nilai estimasi standar deviasi dari residuals

$$\hat{\sigma}_{it} = \sqrt{\frac{RSS}{n-p}}$$

14. Tentukan nilai *internally studentized residual* untuk menentukan adanya *outlier*

$$t_{it} = \frac{e_{it}}{\hat{\sigma}_{it}\sqrt{1-h_{ii}}}$$

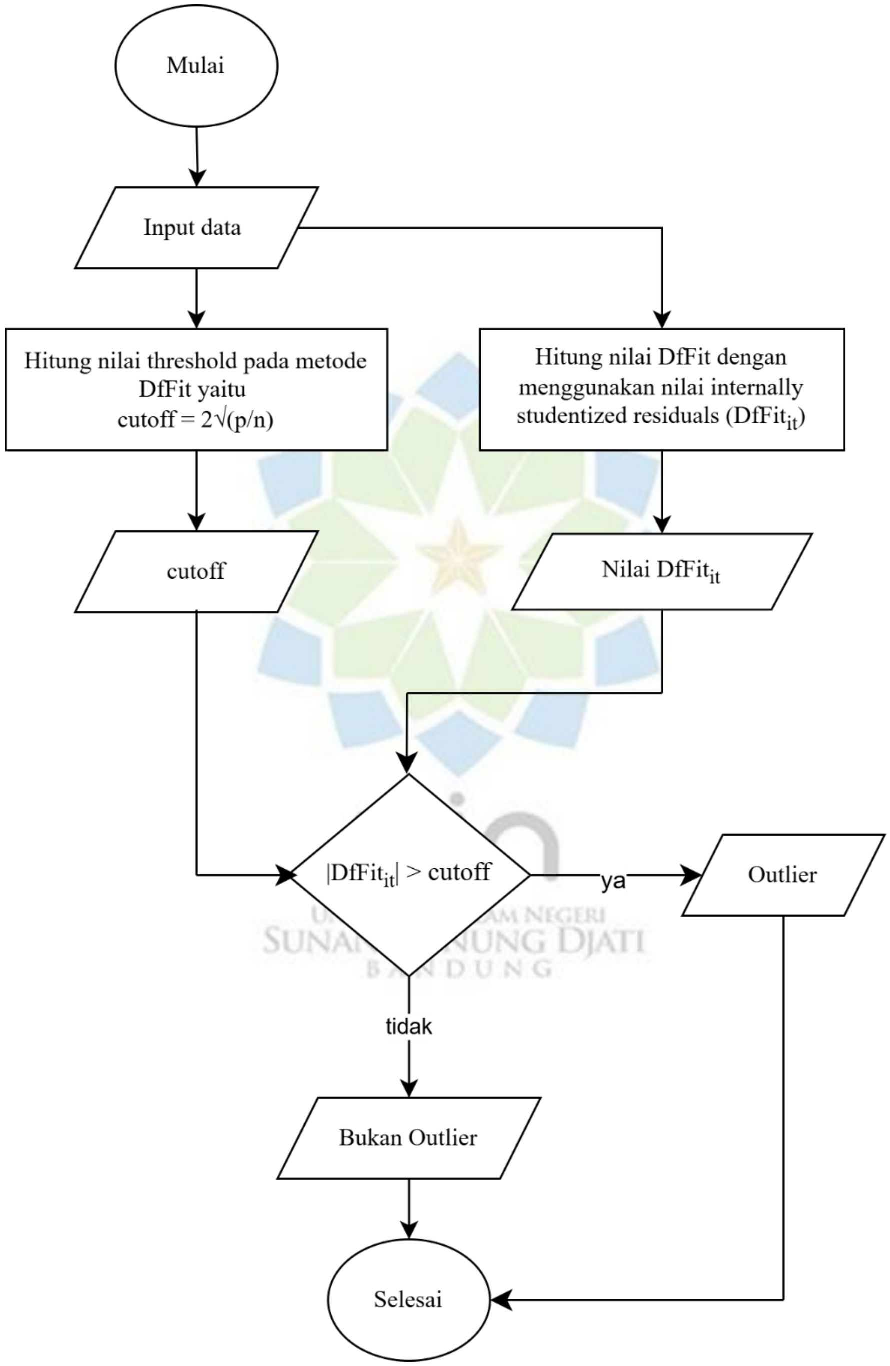
15. Tentukan nilai *DfFit* dengan menggunakan nilai dari *internally studentized residuals*

$$DfFit_{it} = t_{it}\sqrt{\frac{h_{ii}}{1-h_{ii}}} \quad (3.21)$$

Jika nilai $|DfFit_{it}| > 2\sqrt{\frac{p}{n}}$ maka data tersebut dianggap sebagai *outlier*.

16. Dilanjutkan dengan membuat interpretasi hasil dan menarik kesimpulan.

Langkah-langkah pencarian nilai *DfFit* dapat dilihat pada diagram alir yang digambarkan pada Gambar 3.4



Gambar 3.4 *Flowchart* langkah-langkah deteksi *outlier* dengan metode *Difference in Fit (DfFit)*

3.3.4 Metode *Difference in Beta* (*DfBeta*)

Dalam menentukan nilai titik *influence* menggunakan metode *DfBeta* diperlukan langkah-langkah sebagai berikut :

1. Tentukan matriks X_{jit}

$$X_{jit} = \begin{bmatrix} 1 & X_{111} & X_{211} & \cdots & X_{K11} \\ 1 & X_{122} & X_{222} & \cdots & X_{K22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1NT} & X_{2NT} & \cdots & X_{KNT} \end{bmatrix}$$

2. Tentukan matriks transpose dari X_{jit}

$$X_{jit}{}^{T} = \begin{bmatrix} 1 & 1 & & 1 \\ X_{111} & X_{122} & & X_{1NT} \\ X_{211} & X_{222} & \cdots & X_{2NT} \\ \vdots & \vdots & & \vdots \\ X_{K11} & X_{K22} & & X_{KNT} \end{bmatrix}$$

3. Tentukan matriks Y_{it}

$$Y_{it} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{N2} \end{bmatrix}, \ldots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

4. Tentukan nilai matriks $\hat{\beta}_j$

$$\hat{\beta}_j = (X_{jit}{}^{T} X_{jit})^{-1} X_{jit}{}^{T} Y_{it}$$

5. Tetapkan matriks $X_{j(i)t}$

$$X_{j(i)t} = \begin{bmatrix} X_{111} & X_{211} & & X_{K11} \\ X_{121} & X_{221} & & X_{K21} \\ \vdots & \vdots & & \vdots \\ X_{1(i-1)1} & X_{2(i-1)1} & \ldots & X_{K(i-1)1} \\ X_{1(i+1)1} & X_{2(i+1)1} & & X_{K(i+1)1} \\ \vdots & \vdots & & \vdots \\ X_{1N1} & X_{2N1} & & X_{KN1} \end{bmatrix} \cdots \begin{bmatrix} X_{11T} & X_{21T} & & X_{K1T} \\ X_{12T} & X_{22T} & & X_{K2T} \\ \vdots & \vdots & & \vdots \\ X_{1(i-1)T} & X_{2(i-1)T} & \ldots & X_{K(i-1)T} \\ X_{1(i+1)T} & X_{2(i+1)T} & & X_{K(i+1)T} \\ \vdots & \vdots & & \vdots \\ X_{1NT} & X_{2NT} & & X_{KNT} \end{bmatrix}$$

6. Tetapkan matriks $Y_{(i)t}$

$$Y_{(i)t} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{(i-1)1} \\ Y_{(i+1)1} \\ \vdots \\ Y_{N1} \end{bmatrix}, \begin{bmatrix} Y_{12} \\ Y_{22} \\ \vdots \\ Y_{(i-1)2} \\ Y_{(i+1)2} \\ \vdots \\ Y_{N2} \end{bmatrix}, \ldots \begin{bmatrix} Y_{1T} \\ Y_{2T} \\ \vdots \\ Y_{(i-1)T} \\ Y_{(i+1)T} \\ \vdots \\ Y_{NT} \end{bmatrix}$$

7. Tentukan nilai matriks $\hat{\beta}_{j(i)}$

$$\hat{\beta}_{j(i)} = -(X_{j(i)t}{}^{T} X_{j(i)t})^{-1} X_{j(i)t}{}^{T} Y_{(i)t} \tag{3.22}$$

8. Tentukan nilai estimasi varians residual setelah pengamatan ke-i dihapus

$$\hat{\sigma}^2_{(i)t} = \frac{RSS_{(i)}}{n-p-1} \tag{3.23}$$

9. Tentukan nilai varians koefisien regresi $\hat{\beta}_{j(i)}$ di mana pengamatan ke-i dihapus

$$var\left(\hat{\beta}_{j(i)}\right) \approx \hat{\sigma}^2_{(i)t} \left[X_{jit}{}^{T} X_{jit}\right]^{-1}{}_{jj} \tag{3.24}$$

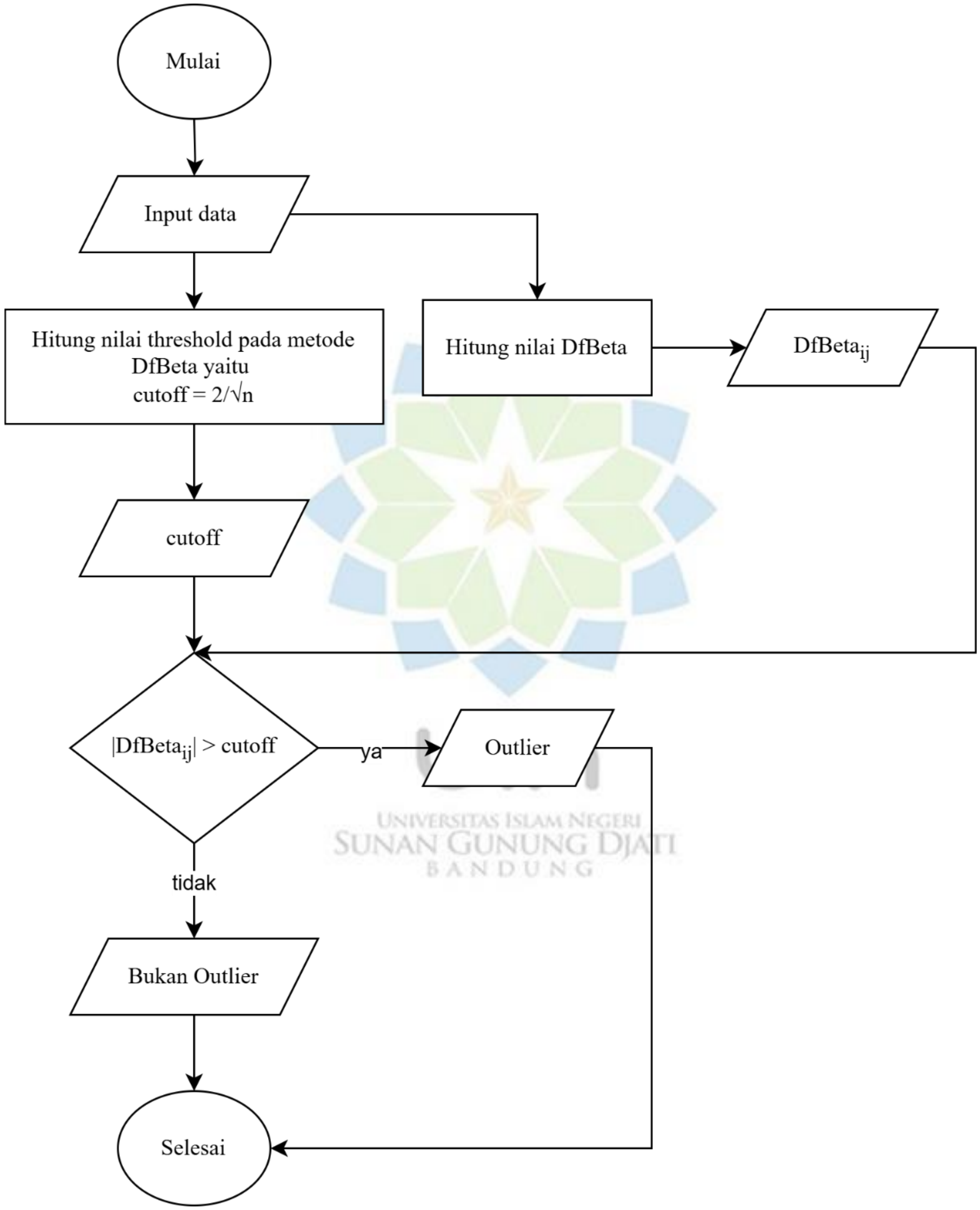
dengan $\left[X_{jit}{}^{T} X_{jit}\right]^{-1}{}_{jj}$ adalah elemen diagonal ke-j dari invers matriks $X_{jit}{}^{T} X_{jit}$

10. Tentukan nilai $DfBeta_{ij}$

$$DfBeta_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{var\left(\hat{\beta}_{j(i)}\right)}} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\hat{\sigma}^2_{(i)t}\sqrt{\left[X_{jit}{}^{T} X_{jit}\right]^{-1}{}_{jj}}} \tag{3.25}$$

11. Identifikasi data dengan cara menentukan jika $DfBeta_{ij} > \frac{2}{\sqrt{n}}$ di mana n adalah banyaknya objek pengamatan, maka data dianggap sebagai *outlier.*
12. Dilanjutkan dengan membuat interpretasi hasil dan menarik kesimpulan.

Langkah-langkah pencarian nilai *DfBeta* dapat dilihat pada diagram alir yang digambarkan pada Gambar 3.5



Gambar 3.5 *Flowchart* langkah-langkah deteksi *outlier* dengan metode *Difference in Beta (DfBeta)*