BAB II

TINJAUAN PUSTAKA

2.1 Kolinieritas

Menurut *Hocking* (1996), pendeteksian adanya kasus kolinieritas dapat dilihat melalui :

a) Koefisien korelasi pearson (r_{jj}) antar variabel indeoenden > 0.95

b) VIF (*Variance Inflation Factors*) > 10

Variance Inflation Factors (VIF) dinyatakan dalam :

$$VIF_j = \frac{Var(\beta_j)}{\sigma^2} = r^{jj} \qquad (2.1)$$

Var (β_j) menunjukkan nilai varian koefisien estimasi variabel predictor ke-j dan σ^2 menunjukkan nilai varian jika variabel-variabel prediktornya ortogonal. Nilai r^{jj} merupakan elemen diagonal ke-j invers matrik korelasi variabel predictor. Nilai VIF_j yang bernilai 1 menunjukkan bahwa variabel-variabel independent tidak saling berkorelasi, jika nilainya lebih dari 10 menunjukkan adanya kolinieritas antara variabel-variabel independen.

2.2 Distribusi *Poisson*

Distribusi *Poisson* merupakan suatu distribusi yang dipergunakan untuk peristiwa yang memiliki probabilitas kejadiannya kecil, dimana kejadian tersebut tergantung pada interval waktu tertentu atau disuatu derah tertentu dengan hasil pengamatan yang berupa variabel diskrit (Rachmah dan Purhadi, 2014). Distribusi *Poisson* memiliki ciri-ciri sebagau berikut (Hassan, 2001) :

a. Banyaknya hasil percobaan yang terjadi dalam suatu interval waktu atau suatu daerah tertentu, tidak bergantung pada banyaknya hasil percobaan yang terjadi pada interval waktu atau daerah lain yang terpisah.

b. Probabilitas terjadinya hasil percobaan selama suatu interval waktu yang singkat atau dalam suatu daerah yang kecil, sebanding dengan panjang interval waktu atau besarnya daerah tersebut dan tidak bergantung pada banyaknya hasil percobaan diluar interval waktu atau daerah tersebut.

c. Probabilitas lebih dari satu hasil percobaan yang terjadi dlam interval waktu yang singkat atau dalam daerah yang kecil dapat diabaikan.

Distribusi *Poisson* digunakan dalam :

a. Menghitung probabilitas terjadinya peristiwa menurut satuan waktu ruang atau isi dan luas.

b. Menghitung distribusi binomial apabila n-besar ($n \geq 30$) dan p relative kecil ($p < 0,1$)

Rumus dari Distribusi *Poisson* adalah :

$$P(X) = \frac{e^{\lambda}\lambda^{x}}{x!} \tag{2.2}$$

Dimana :

λ = np

n = banyaknya amatan

p = probabilitas sukses

x = variabel random doskrit

e = bilangan trasional (2,71828)

2.3 Regresi *Poisson*

Regresi *poisson* merupakan analisis regresi nonlinier dari distribusi *poisson*, dimana analisis ini sangat cocok digunakan dalam menganalisis data diskrit (*count*). Model Regresi *Poisson* merupakan *Generalized Linier Model* (GLM) yang data respon diasumsikan berdistribusi *Poisson*. Model Regresi *Poisson* diberikan sebagai berikut.

$$f(y,\mu) = \frac{e^{-\mu}\mu^{y}}{y!}; y = 0,1,2,\ldots,n \qquad (2.3)$$

Dengan μ merupakan rata-rata variabel respon yang berdistribusi *Poisson* dimana nilai rata-rata dan varian dari Y mempunyai nilai lebih dari 0. Persamaan model regresi *Poisson* dapat ditulis sebagai berikut.

$$\mu_i = exp\,(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}) \qquad (2.4)$$

Dengan μ_i merupakan rata-rata jumlah kejadian yang terjadi dalam interval waktu tertentu.

Pada regresi *poisson* diasumsikan variabel respon (Y) berdistribusi *poisson* dan tidak terjadi multikolinearitas diantara masing-masing variabel prediktor (X). Dalam regresi *poisson* terdapat asumsi yang harus dipenuhi yaitu variabel respon (Y) diskrit dan asumsi *equidispersi*. *Equidispersi* yaitu nilai rata-rata sama dengan nilai varian atau $V(Y) = E(Y) = \mu$.

2.4 *Overdispersi*

Pada model Regresi *Poisson* terdapat asumsi yang mendasari dalam melakukan analisis Regresi *Poisson* yaitu dalam variabel terikat harus terjadi *equidispersi* (rata-rata sama dengan variansi), yaitu *E(Y)* = *Var(Y)*, (Irwan dan Sari, 2013). Namun terkadang terjadi kasus *overdispersi* dan *underdispersi*. *Overdispersi* yaitu terjadi jika kasus nilai *varians* lebih besar dari *mean*nya, sedangkan *underdispersi* terjadi jika nilai *varians* lebih kecil dari *mean*nya (Camelia *et al*., 2016). Jika pada data diskrit terjadi *overdispersi* teteapi tetap digunakan model regresi *poisson*, maka estimasi parameter koefisien regresinya tetap konsisten tetapi tidak efisien karena berdampak pada nilai standar error yang tinggi. Hal-hal penyebab *overdispersi* antara lain :

a. Terjadi korelasi antara pengamatan
b. Asumsi *equidispersi* regresi *poisson* tidak terpenuhi
c. Terdapat *excess zero*
d. Terdapat outlier

Ada atau tidaknya *overdispersi* dapat dilihat dari nilai *Deviance* atau *Pearson Chi-square* yang dibagi dengan derajat bebasnya. Apabila nilai *Pearson Chi-Square* dibagi dengan derajat bebas lebih besar daripada 1, ini menunjukkan nilai variansi yang lebih besar daripada rataan yang artinya telah terjadi *overdispersi* (Khoshgoftaar, 2014). Salah satu model yang dapat digunakan untuk mengatasi masalah *overdispersi* adalah dengan model *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial*.

2.5 *Excess Zeros*

Salah satu permasalahan pada regresi *poisson* yaitu nilai nol yang berlebih (*Excess zeros*). Pada variabel respon pada data diskrit mungkin data yang bernilai kosong/nol. Akan tetapi, dalam banyak kasus, kosong memiliki arti penting pada penelitian yang bersangkutan. Jika nilai nol memiliki arti penting dalam data diskrit maka data tersebut harus dimasukkan dalam analisis. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai lebih besar dari data diskrit artinya, *Excess zeros* merupakan salah satu penyebab terjadinya *overdispersi* (Ariawan, 2012)

2.6 Metode Maksimum Likelihood

Misalkan X_1, X_2, ... X_n adalah sampel random dari populasi dengan densitas $f(x;\theta)$.dengan $\theta = (\theta_1 \ldots, \theta_p)^T$ maka fungsi likelihood didefinisikan sebagai fungsi densitas bersama dari $x_1, x_2, \ldots, x_n$, sehingga

$$L(\theta) = f(x1, x2, \ldots xn; \theta) = f(x1;\theta).f(x2;\theta)\ldots f(xn;\theta) = (x+a)^n = \Pi_{i=1}^{n} f(xi;\theta) \tag{2.5}$$

Estimator maksimum likelihood adalah nilai $\hat{\theta}$ yang memaksimalkan fungsi likelihood $L(\boldsymbol{\theta})$. Untuk memperoleh nilai $\hat{\theta}$ yang memaksimumkan $L(\boldsymbol{\theta})$ harus diderivatifkan dengan langkah-langkah sebagai berikut :

1. Nilai $\hat{\theta}$ diperoleh dari derivative pertama

$$\frac{\partial}{\partial \theta_j} L(\theta) = 0 \; dengan \; j = 1,2,3, \ldots p \tag{2.6}$$

2. Nilai $\hat{\theta}$ dikatakan memaksimumkan $L(\theta)$ jika

$$\frac{\partial^2}{\partial \theta_j} L(\theta) = \Big|_{\theta=\hat{\theta}} < 0 \;\; dengan \; j = 1,2,3, \ldots p \tag{2.7}$$

Selain memaksimumkan fungsi likelihood, nilai $\hat{\theta}$ juga dapat diperoleh dengan memaksimumkan log natural-likelihood (*ln L(θ)*). Dalam banyak kasus dengan diferensiasi digunakan, akan lebih mudah bekerja pada logaritma natural yang dinotasikan dengan *l(θ)* = *ln L(θ)*. Untuk memperoleh nilai $\hat{\theta}$ yang memaksimumkan *ln L(θ)* dapat dilakukan dengan langkah-langkah yang sama seperti dalam memperoleh nilai $\hat{\theta}$ yang memaksimumkan *L(θ)*.

2.7 Regresi *Zero-Inflated Poisson*

Menurut *Jansakul,et.al.,(2001)* bahwa salah satu penyebab terjadinya overdisersi adalah lebih banyaknya observasi yang bernilai nol dari pada yang ditaksir untuk model regresi *Poisson*. Salah satu metode analisis yang diusulkan untuk kasus seperti ini adalah model regresi *Zero-Inflated Poisson (ZIP).* Jika y_i adalah variabel random independen yang mempunyai distribusi ZIP, observasi diduga muncul dalam dua cara yang sesuai untuk *state* yang terpisah. *State* pertama terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol, sementara *state* kedua terjadi dengan probabilitas $(1-\omega_i)$ dan berdistribusi *Poisson* dengan mean μ_i. Proses dua *state* ini memberikan distribusi campuran dua komponen dengan fungdi probabilitas sebagai berikut :

$$Pr(Y_i = y_i) = \left\{ \begin{matrix} \omega_i + (1-\omega_i)e^{-\mu_i}, y_i = 0 \\ (1-\omega_i)\frac{e^{-\mu_i}\mu_i^{\mu_i}}{y_i!}, y_i = 1,2, \ldots 0 \leq \omega_i \leq 1 \end{matrix} \right\} \tag{2.8}$$

Mean dan Varians masing-masing didefinisikan sebagai berikut :

$$E(Y) = (1-\omega)\mu, \tag{2.9}$$

$$Var(Y) = \mu(1-\omega)(1+\mu\omega) \tag{2.10}$$

Lambert dalam *Jansakul,et.al.,(2001)* menunjukkan model golongan untuk μ dan ω sebagai berikut :

$$log(\mu) = X\beta \ dan \ logit\ (\omega) = log\left(\frac{\omega}{1-\omega}\right) = X\gamma \tag{2.11}$$

Dimana X adalah matriks variabel prediktor sedangkan β dan γ adalah parameter yang akan ditaksir.

2.7.1 Estimasi Parameter Model Regresi *Zero-Inflated Poisson*

Khoshgoftaar,et.al.,(2004) menjelaskan bahwa estimasi parameter regresi ZIP dengan menggunakan metode Maximum Likelihood Estimation (MLE). Fungsi Likelihood ZIP adalah sebagai berikut :

$$L(\beta,\gamma|y_i,x_i) = \left\{\begin{array}{l} \prod_{i=1}^{n} \frac{exp(x_i^T\gamma)+exp\left(-exp(x_i^T\beta)\right)}{1+exp(x_i^T\gamma)}, y_i = 0 \\ \prod_{i=1}^{n} \frac{\frac{1}{1+exp\left(x_i^T(\gamma)\right)}\left(exp\left((-exp(x_i^T\beta)+(x_i^T\beta)y_i)\right)\right)}{y_i!}, y_i > 0 \end{array}\right\} \tag{2.12}$$

dimana x_i adalah variabel prediktor, y_i adalah variabel respon, serta β dan γ adalah parameter yang akan ditaksir.

Fungsi Log-Likelihood gabungan untuk model regresi ZIP diberikan oleh :

$$\begin{aligned} lnL(\beta,\gamma|y_i,x_i) = & \sum_{\substack{i=1 \\ y_i=0}}^{n} ln\left(exp(x_i'\gamma)\right) - \sum_{i=1}^{n} ln\left(1+exp(x_i'\gamma)\right) + \\ & \sum_{\substack{i=1 \\ y_i>0}}^{n} \left((y_i x_i^T\beta) - exp(x_i^T\beta)\right) - \\ & \sum_{\substack{i=1 \\ y_{i>0}}}^{n} ln(y_i)! \end{aligned} \tag{2.13}$$

Estimasi maximum likelihood untuk β dan γ dapat di peroleh dengan menggunakan pendekatan standart untuk model campuran, yaitu Algoritma EM. Algoritma EM memberikan prosedur sederhana yang dapat diimplementasi dalam software standar, atau metode estimasi langsung seperti metode Newton-Raphson.

2.8 Regresi *Zero-Inflated Negative Binomial* (ZNIB)

Hilbe (2011) mengemukakan bahwa, regresi *Zero-Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran *poisson* gamma.

Jika Yi adalah variabel random independen yang diskrit dengan i = 1,2,3,…,n , nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Keadaan pertama disebut *zero state* terjadi dengan probabilitas *pi* dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Negative Binomial state* terjadi dengan probabilitas (1 - *pi*) dan berdistribusi Binomial Negatif dengan mean μ , dengan 0 ≤ *pi* ≤ 1. Proses dua keadaan ini dengan variabel Yi memberikan distribusi campuran dua komponen dan didapat fungsi probabilitas sebagai berikut :

$$P(Y_i = y_i) = \begin{cases} p_i + (1-p_i)\left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}} & , y_i = 0 \\ (1-p_i)\frac{\Gamma(y_i+\frac{1}{k})}{\Gamma(\frac{1}{k})\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}}\left(\frac{1}{1+k\mu_i}\right)^{y_i} & , y_i = 1,2,\ldots, \end{cases} \quad (2.14)$$

Dimana μ_i dan k adalah ukuran parameter yang masing-masing distribusi campurannya adalah Zero-Inflated Negative Binomial. Mean dan Varians masing-masing didefinisikan sebagai berikut:

$$E(Y_i) = (1-p_i)\mu_i, \quad (2.15)$$

$$Var(Y_i) = (1-p_i)\mu_i(1+k\mu_i+p_i\mu_i). \quad (2.16)$$

Diasumsikan bahwa parameter μi dan pi masing-masing bergantung pada variabel xi ,sehingga model dari regresi ZINB dibagi menjadi dua komponen yaitu model data diskret untuk μi dan model *Zero-Inflation* untuk pi yaitu:

$$ln(\mu_i) = x_i^T\beta, \qquad \mu_i \geq 0, i = 1,\ldots,n. \quad (2.17)$$

$$logit(p_i) = ln\left(\frac{pi}{1-pi}\right) = z_i^T\gamma, \quad ,0 \leq pi \leq 1, i = 1,..,n. \tag{2.8}$$

x_i adalah matriks variabel yang memuat himpunan-himpunan yang berada dari factor eksperimen yang berhubungan dengan peluang pada mean *Negative Binomial* dan *Negative Binomial State*, sedangkan β dan γ adalah parameters regresi yang akan ditaksir.

2.8.1 Estimasi Parameter Regresi *Zero-Inflated Negative Binomial* (ZNIB)

Bayu Ariawan, Suparti, dan Sudarno (2012) menyatakan bahwa estimasi parameter model regresi ZINB menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan prosedur Algoritma EM (*Expectation Maximization*) dan Newton Rhapson. Metode ini biasanya digunakan untuk menduga parameter suatu model yang diketahui fungsi densitasnya. Sehingga fungsi *log-likelihood* dari fungsi probabilitas ZINB adalah :

$$lnL(\theta|y_i) = \begin{cases} \sum_{i=1}^{n} ln\left\{e^{x_i^T\gamma} + \left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\right\} - \sum_{i=1}^{n} ln\left[1 + ke^{x_i^T\gamma}\right] & ,y_i = 0 \\ -\sum_{i=1}^{n} ln\left[1 + ke^{x_i^T\gamma}\right] + \sum_{i=1}^{n} ln\left[\Gamma\left(\frac{1}{k} + y_i\right)\right] - \sum_{i=1}^{n} ln\left[\Gamma(y_i + 1)\right] - \sum_{i=1}^{n} ln\left[\Gamma\left(\frac{1}{k}\right)\right] & \\ +y_i \sum_{i=1}^{n} ln\left(\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right)^{y_i} + \frac{1}{k}\sum_{i=1}^{n} ln\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}} & ,y_i = 1,2,\ldots \end{cases} \tag{2.9}$$

Karena fungsi log-likelihoodnya tidak linier jika tidak digunakan nilai awal yang bagus, sehingga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa. Sehingga digunakanlah algoritma EM (*Expectation Maximization*) (Garay & Hashimoto, 2011).

Misalkan variabel y_i(i = 1,2,3…,n) berkaitan dengan vector variabel indicator W = (w_1,….,w_n)T yaitu :

$$W_i = \begin{cases} 1, jika\ y_i\ berasal\ dari\ zero\ state \\ 0, jika\ y_i\ berasal\ dari\ Negative\ Binomial\ State \end{cases}$$

Permasalahan disini pendefisinian *Negative Binomial State*, yang bisa berasal dari 0 maupun dari *zero state* dan *negative binomial state*. Permasalahan ini dapat diatasi menggunakan algoritma EM. Algoritma EM terdiri dari dua tahap yaitu tahap Ekspetasi (tahap perhitungan ekspetasi dari fungsi ln *Likelihood*) dan tahap Maksimalisasi, perhitungan untuk mencari estimasi parameter yang memaksimumkan fungsi ln *likelihood* hasil dari tahap ekspetasi sebelumnya.

2.9 Pengujian kesesuaian Model

Menurut Zamzani & Ismail (2013) pengujian kesesuaian model regresi ZIP ZINB adalah dengan menggunakan *Likelihood Ratio (LR) test.* Hipotesis untuk pengujian kesesuaian model adalah sebagai berikut :

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = \gamma_1 = \gamma_2 = \cdots \gamma_k = 0$$

$$H_1: paling\ sedikit\ ada\ satu\ \beta_r \neq 0\ atau\ \gamma_r \neq 0, 0 < r < k-1$$

dimana k+1 adalah jumlah parameter, β_r adalah parameter model log ke-r, dan γ_r adalah parameter model logit ke-r. *Lestari* (2008) telah melakukan perhitungan statistik uji untuk pengujian kesesuaian model sebagai berikut :

$$G = -2\ln\left[\frac{L(y;\hat{\omega})}{L(y;\hat{\Omega})}\right]$$

$$= \left(2\sum_{i=1}^{n}\left(z_i x_i^T\hat{\gamma} - ln\left(1 + exp\left(x_i^T\hat{\gamma}\right)\right) + 2\sum_{i=1}^{n}(1-z_i)\left(y_i x_i^T\hat{\beta} - exp\left(x_i^T\hat{\beta}\right)\right)\right) - 2\sum_{i=1}^{n} z_i\hat{\gamma}_0 - ln\left(1 + x_i^T\hat{\gamma}_0\right) + 2\sum_{i=1}^{n}(1- z_i)\left(y_i\hat{\beta}_0 - exp\left(\hat{\beta}_0\right)\right)\right) \quad (2.20)$$

Pengujian parameter secara individu ada dua, yaitu pengujian parameter model log dan pengujian parameter model logit. Statistik uji untuk pengujian parameter model log secara individu adalah sebagai berikut (*Lestari,2008*) :

$$G = -2\ln\left[\frac{L(y;\widehat{\omega})}{L(y;\widehat{\Omega})}\right]$$

$$= \left(2\sum_{i=1}^{n}\left(z_i x_i^T \hat{y} - ln\left(1 + exp(x_i^T \hat{\gamma}\right)\right) + 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T \hat{\beta} - exp\left(x_i^T \hat{\beta}\right)\right)\right) - 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T \hat{\beta}_i - exp\left(x_i^T \hat{\beta}_i\right)\right)\Big) \quad (2.21)$$

Statistik uji yang digunakan untuk pengujian parameter model logit adalah sebagai berikut *(Lestari,2008)* :

$$G = -2\ln\left[\frac{L(y;\widehat{\omega})}{L(y;\widehat{\Omega})}\right]$$

$$= 2\sum_{i=1}^{n}\left(z_i x_i^T \hat{y} - ln\left(1 + exp(x_i^T \hat{\gamma}\right)\right) + 2\sum_{i=1}^{n}(1 - z_i)\left(y_i x_i^T \hat{\beta} - exp\left(x_i^T \hat{\beta}\right)\right)\Big) - 2\sum_{i=1}^{n}(1 - z_i) ln(y_i)! - 2\sum_{i=1}^{n}\left(z_i \hat{\gamma}_0 - ln\left(1 + exp(\hat{\gamma}_0)\right)\right) \quad (2.10)$$

Daerah penolakan untuk ketiga pengujian diatas adalah tolak H_0 pada taraf signifikansi α jika $G_{hitung} > \mathcal{X}^2_{(v,a)}$, dimana v adalah derajat bebas.

2.9.1 Pengujian Signifikansi Parameter

Myers et al. (2010;218) menjelaskan bahwa pengujian signifikansi menggunakan uji *wald* dibagi menjadi 2 model yaitu :

2.9.1.1 Pengujian Signifikansi Parameter $\boldsymbol{\beta}$

Pengujian signifikansi parameter model $\ln(\mu_i) = x_i^T \beta$ dengan $i = 1, \ldots, n$ untuk model regresi ZINB dan parameter model $log(\mu) = X\beta$ untuk model regresi ZIP.

Hipotesis :

H_0: $\beta_i = 0$ (tidak ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

H_1: $\beta_i \neq 0$ (ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

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

Statistika Uji :

$$W_i = \left(\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}\right)^2 \tag{2.11}$$

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_i > \chi^2_{(a,1)}$ pada taraf alfa. Penolakan H_0 menjelaskan bahwa X_i memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi.

2.9.1.2 Pengujian Signifikansi Parameter γ

Pengujian signifikansi parameter model $\ln\left(\frac{pi}{1-pi}\right) = Z_i^T\gamma$, dengan $i = 1, \dots, n$ untuk regresi ZINB dan parameter model $logit\ (\omega) = log\left(\frac{\omega}{1-\omega}\right) = X\gamma$ untuk model regresi ZIP.

Hipotesis :

$H_0: \gamma_i = 0$ (tidak ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

$H_1: \gamma_i \neq 0$ (ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

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

Statistika Uji :

$$W_i = \left(\frac{\hat{\gamma}_i}{SE(\hat{\gamma}_i)}\right)^2 \tag{2.24}$$

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_i > \chi^2_{(a,1)}$ pada taraf alfa. Penolakan H_0 menjelaskan bahwa X_i memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi.

2.10 Pemilihan Model Terbaik

Metode *Akaike's Information Criterion* (AIC) merupakan salah satu metode yang digunakan untuk memilih model terbaik dari Regresi ZIP dan ZINB. Nilai AIC adalah sebagai berikut *(Dalrymple,et.al.,(2001)* :

$$AIC = G + (k + 1) \tag{2.25}$$

dimana G adalah statistik uji kesesuaian model, dan $k + 1$ adalah jumlah parameter. Model terbaik Regresi *Poisson* dan ZIP adalah model dengan nilai AIC terkecil.

2.11 Tanah Longsor

Menurut Rudiyanto (2016), bencana alam longsoran tanah yang banyak terjadi di Indonesia merupakan salah satu jenis gerakan massa tanah (*soil mass movement*) pada lereng-lereng alam. Apabila massa yang bergerak ini didominasi oleh massa tanah dan gerakannya melalui suatu bidang pada lereng, baik berupa bidang miring ataupun lengkung, maka proses pergerakan tersebut disebut sebagai longsoran tanah. Terjadinya bencana alam gerakan tanah ataupun longsoran terutama karena gangguan secara alamiah pada kestabilan tanah dan atau batuan penyusun lereng, baik yang bersifat alamiah maupun non alamiah. Gerakan tanah ataupun longsoran akan dikategorikan sebagai bencana apabila terjadi pada daerah yang dihuni oleh manusia atau pada daerah tempat kegiatan manusia. Jadi aspek kehadiran manusia atau terpengaruhnya aktivitas manusia sangat penting dalam menetapkan apakah suatu gerakan tanah atau longsoran dianggap sebagai bencana atau tidak.

2.11.1 Ketinggian Wilayah

Faktor penyebab tanah longsor juga merupakan salah satu penentu kerawanan longsor. Ketinggian suatu daerah baik itu pada dataran rendah, dataran tinggi bahkan pegunungan dapat menentukan besarnya kekuatan tanah yang terjatuh apabila terjadi longsor. Semakin tinggi suatu tempat semakin besar kekuatan tanah yang terjatuh karena adanya pengaruh gravitasi (Noor, 2005).

2.11.2 Kepadatan Penduduk

Kepadatan penduduk adalah perbandingan antara jumlah penduduk dengan luas wilayah yang didiami oleh penduduk di suatu wilayah yang dinyatakan dalam satuan jiwa/km2. Penilaian kepadatan penduduk berpedoman pada kriteria Peraturan Kepala BNPB Nomor 02 (2012 : 29) (Ni'mah, 2017).

2.11.3 Penggunaan Lahan

Penggunaan lahan mempunyai pengaruh besar terhadap kondisi air tanah, hal ini akan mempengaruhi kondisi tanah dan batuan yang pada akhrinya juga akan mempengaruhi keseimbangan lereng. Pengaruhnya dapat bersifat memperbesar atau memperkecil kekuatan geser tanah pembentuk lereng. Selanjutnya mengenai harkat penggunaan lahan di daerah penelitian mendasarkan pada klasifikasi penggunaan lahan dengan sedikit modifikasi sesuai dengan kondisi daerah penelitian (Rudiyanto, 2016).

2.11.4 Curah Hujan

Curah hujan merupakan salah satu faktor penentu tingkat potensi bahaya longsor di daerah penelitian. semakin tinggi nilai curah hujannya, maka sudah dapat dipastikan bahwa wilayah tersebut merupakan wilayah yang mempunyai potensi tertinggi terjadi bencana tanah longsor. Untuk lebih lengkapnya mengenai klasifikasi curah (Rudiyanto, 2016).