BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi Logistik

Regesi logistik adalah salah satu metode yang dapat digunakan untuk mencari hubungan variabel respon yang bersifat dikotomus (bersekala nominal atau ordinal dengan dua kategori) atau bersifat polikotomus (yang mempunyai skala nominal atau ordinal lebih dari dua kategori) dengan satu atau lebih variabel prediktor yang bersifat kontinu atau kategorik. Perbedaan regresi linier sederhana dan regresi logistik terletak pada variabel respon dimana respon pada regresi logistik adalah berupa kategorik Agresti(2002). Pada regresi logistik, untuk variabel responnya dianggap hanya mempunyai dua nilai yang mungkin, yaitu 0 (gagal) atau 1 (sukses), sehingga variabelrespon (y) tersebut mengikuti distribusi Bernoulli dengan fungsi peluang distribusi) sebagai berikut:

$$P(\, Y = y\,) = \pi^{y}(1-\pi)^{1-y};\, y = 0, 1 \qquad (1)$$

Dimana jika y = 0 maka P(Y = 0) = 1 – π $dan\ jika\ \ y = 1\, maka$ P(Y = 1) = π

Bentuk umum fungsi hubungan yang digunakan adalah fungsi hubung logit, maka distribusi peluang yang digunakan adalah fungsi logistik

$$\pi(x) = \frac{e^{\beta_0+\beta_1+\beta_2+\beta_2+...+\beta_p+\beta_p}}{1+e^{\beta_0+\beta_1+\beta_2+\beta_2+...+\beta_p+\beta_p}} \qquad (2)$$

Untuk mempermudah dalam pendugaan parameter regresi, maka (x) ditransformasi dengan menggunakan tranformasi logit sehingga diperoleh bentuk sebagai berikut (Fotheringham, et al.) :

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p \qquad (3)$$

Bentuk logit g(x) ini merupakan model logit, fungsi linear dalam parameter-parameternya, dan berada dalam jarak antara -∞sampai +∞ tergantung dari variabel X.

2.2 Regresi Logistik Multinomial

Regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel dependen mempunyai skala yang bersifat polichotomous atau multinomial. Skala multinomial adalah suatu pengukuran yang dikategorikan menjadi lebih dari dua kategori. Metode yang digunakan dalam penelitian ini adalah regresi logistik dengan variabel dependen berskala nominal dengan tiga kategori. Mengacu pada regresi logistik trichotomous[5] untuk model regresi dengan variabel dependen berskala nominal tiga kategori digunakan kategori variabel hasil Y dikoding 0,1, dan 2. Variabel Y terparameterisasi menjadi dua fungsi logit. Sebelumnya perlu ditentukan kategori hasil mana yang digunakan untuk membandingkan.Pada umumnya digunakan Y=0 sebagai pembanding. Untuk membentuk fungsi logit, akan dibandingan Y=1 dan Y=2, terhadap Y=0.

Bentuk model regresi logistik dengan p variabel prediktor(Yudissanta dan Ratna,2012)

Dengan menggunakan transformasi logit akan didapatkan dua fungsi logit,

$$g_1(x) = ln\left[\frac{p(y=2|x)}{P(Y=0|x)}\right] = x\beta \qquad (4)$$

$$g_2(x) = ln\left[\frac{p(y=2|x)}{P(Y=0|x)}\right] = x\beta \qquad (5)$$

Berdasarkan kedua fungsi logit tersebut maka didapatkan model regresi logistik trichotomous sebagai berikut

$$\pi_0(x) = \frac{1}{1+expg_1(x)+expg_2(x)} \qquad (6)$$

$$\pi_1(x) = f(x) = \frac{\exp g_1(x)}{1+expg_1(x)+expg_2(x)} \qquad (7)$$

$$\pi_2(x) = f(x) = \frac{\exp g_2(x)}{1+expg_1(x)+expg_2(x)} \qquad (8)$$

Dengan P(Y=j|x)π_j(x) untuk j= 0,1,2.

Pengujian regresi logistik dilakukan dengan dua cara yaitu pengujian parameter secara serentak dilakukan untuk mengetahui signifikansi parameter β terhadap variabel respon dengan menggunakan statistik uji G dan pengujian parameter secara parsial dilakukan untuk mengetahui signifikansi parameter β terhadap variabel respon secara parsial dengan meggunakan statistik *uji Wald* (Hosmer dan Lemeshow, 2000)

2.3 Pengujian Parameter Model Regresi Logistik

Pengujian parameter model regresi logistik dilakukan dengan menggunakan *metode Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis pada uji serentak sebagai berikut

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : paling tidak terdapat satu $\beta_k \neq 0$; k = 0, 1, 2, … p

Statistik uji :

$$G^2 = \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n} \hat{\pi i}^{yi} (1-\hat{\pi})^{1-yi}} \right] \quad (9)$$

Daerah penolakan : tolak Ho jika $G^2 > x^2\ (v, \alpha)$ dengan v adalah derajat bebas banyaknya variabel prediktor. Pengujian parameter secara parsial digunakan untuk mengetahui parameter yang berpengaruh signifikan terhadap model. Hipotesis yang digunakan adalah sebagai berikut

H_0 : $\beta_k = 0$

H_1 : $\beta_k \neq 0$; k = 0, 1, 2, …. P

Statistik uji W $= \frac{\acute{\beta}_k}{SE\ (\beta_k)}\ atau\ \frac{\acute{\beta}_k 2}{SE\ (\beta_k)2}$ (10)

Daerah penolakan : tolak H_0 jika$|W| > Z_{\alpha/2}$ atau tolak H_0 Jika $W^2 > x^2(v,\alpha)$ dengan v adalah derajat bebas banyaknya variabel prediktor.

2.4 Model *Geographically Weighted Regression*

GWR merupakan salah satu metode yang cukup efektif untuk mengestimasi data yang memiliki *spatial heterogenity* (ketidak seragaman dalam spasial)(Leung, Mei & Zhang, 2000). Pada analisis spasial, seringkali data digambarkan dalam suatu unit geografis tertentu dan diestimasi menggunakan satu persamaan regresi global. Hal tersebut akan menghasilkan estimasi parameter lokal, dimana masing-masing area penelitian akan memiliki parameter yang berbeda (Brunsdon, Fotheringham & Charlton, 1998). Pada model GWR, diasumsikan bahwa masing-masing lokasi pengamatan dalam satu wilayah regional memiliki koordinat spasial. Koordinat spasial pada lokasi pengamatan ke-i dilambangkan dengan (u_i ,v_i) Persamaan umum GWR (Fotheringham, Brunsdon & Charlton, 2002) adalah sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{i=1}^{4} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i, i = 1, 2, ..., n \qquad (11)$$

dimana y_i adalah nilai variabel dependen pada pengamatan ke-i x_{ij} adalah nilai variabel independen ke-j pada pengamatan ke-I $\beta_0(u_i, v_i)$, adalah konstanta/intercept pada pengamatan ke-i, $\beta_j(u_i, v_i)$ adalah nilai fungsi variabel independen x_j pada pengamatan ke-i, p adalah jumlah variabel independen, (u_i, v_i) , adalah titik kordinat lokasi pengamatan ke-i, dan ε adalah random error yang diasumsikan berdistribusi $N\left(0, \sigma^2 I\right)$ dengan $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ dan I adalah matriks identitas.

Estimasi parameter GWR menggunakan metode *Weighted Least Square* (WLS) yaitu dengan memberikan pembobot yang bebeda untuk setiap lokasi pengamatan [10]. Estimasi parameter pada GWR dapat ditulis sebagai berikut:

$$\beta(u_i,v_i)=(X^TW(u_i,v_i)X)^{-1}X^TW(u_i,v_i)y \tag{12}$$

dimana $\mathbf{W}(u_i,v_i)$, adalah matriks pembobot berukuran n x n. Fungsi pembobot untuk memberikan hasil penduga parameter yang berbeda pada lokasi yang berbeda. Penelitian ini menggunakan fungsi kernel gaussian (Fotheringham,2002) dengan persamaan sebagai berikut:

$$\boldsymbol{W}_J(u_iv_i\exp(-\tfrac{1}{2}(d_{ij}/h)^2 \tag{13}$$

2.5 Model *Geographically Weighted Logistic Regression* (GWLR)

Metode GWLR adalah metode non parametrik pada regresi yang mempertimbangkan faktor spasial. GWLR merupakan pengembangan metode GWR (*Geographically Weighted Regression*) dengan data respon yang berbentuk kategorik. Dalam model GWLR, variabel respon y diprediksi dengan variabel independen yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati (Fotheringham, et al). Model matematis dari metode GWLR dijelaskan sebagai berikut :

$$\pi(x_i)=\frac{\exp(\sum_{k=0}^{4}\beta_k(u_i,v_i)x_{ik})}{1+exp(\sum_{k=0}^{p}\beta_x(u_{i,}v_i)x_{ik})} \tag{14}$$

$$g(xi)=\ln\left[\frac{\pi(xi)}{1-\pi(xi)}\right]=\beta o(ui,vi)+\beta 1(ui,vi)xi1+...+\beta k(ui,vi)xik \tag{15}$$

2.5.1 Penaksiran Parameter Model GWLR

Penaksiran parameter dalam model GWLR adalah menggunakan *Maximum Likelihood Estimation* (MLE). Langkah pertama dari metode tersebut adalah dengan membentuk fungsi likelihood dengan variabel respon berdistribusi Bernoulli.

$$L(\beta(u_i,v_i)) = \left\{ \prod_{i=1}^{n} \left[1 + \exp \sum_{k=0}^{p} \beta_k(u_i,v_i) x_{ik} \right]^{-1} \right\} \exp \left[\sum_{k=0}^{p} \left(\sum_{k=0}^{n} \beta_k(y_i, x_{ik} \right) \beta_k(y_i, x_{ik} \right] \tag{16}$$

Fungsi ln likelihoodnya menjadi :

$$L\left(\beta(u_i,v_i)\right) = \sum_{i=1}^{p} \left(\sum_{j=i}^{n} y_j x_{jk} \right) \beta_X(u_i,v_i) - \sum_{i=1}^{n} \ln \left\{ 1 + \exp \left(\sum_{i=1}^{p} \beta_X(u_i,v_i) x_{jk} \right) \right\} \tag{17}$$

Taksiran parameter yang dihasilkan dari MLE berbentuk implisit sehingga digunakan metode *Newton Raphson Iteratively Reweighted Least Square*(IRLS) Secara umum persamaan untuk iterasi Newton Rhapson adalah :

$$\beta^{(t+1)}(\boldsymbol{u_i}, \boldsymbol{v_i}) = \boldsymbol{\beta}^{(t)} - \boldsymbol{H}^{(\boldsymbol{T})-\boldsymbol{1}}(\boldsymbol{\beta}^{\boldsymbol{T}}(\boldsymbol{u_i}, \boldsymbol{v_i}))g^{(t)}\left(\boldsymbol{\beta}^{(t)}(\boldsymbol{u_i}, \boldsymbol{v_i})\right) \tag{18}$$

dimana g merupakan turunan pertama dari fungsi ln likelihood dan $H^{(t)}\left(\beta^{(t)}(u_i, v_i)\right)$ adalah matriks Hessian dengan elemen-elemennya adalah

elemennya adalah $h_{kk} = \dfrac{\partial^2 \ln L^*(\beta(u_i,v_i))}{\partial \beta_k(u_i,v_i) \partial \beta_{k^*}(u_i,v_i)}$ (19)

2.5.2 Pengujian Parameter Model GWLR

Pengujian secara parsial dilakukan untuk mengetahui variabel mana saja yang berpengaruh signifikan terhadap variabel responnya dengan hipotesis yang diuji sebagai berikut:

H_0 : $\beta_k = \beta_k(u_i , v_i) = 0$

H_1 : $\beta_k (u_i , v_i) \neq 0$

Statistika uji yang digunakan adalah :

$$z = \frac{\beta k(u_i v_i)}{se\left(\beta_k (u_i v_i)\right)} \tag{20}$$

Kriteria pengujiannya adalah tolak Ho jika $|z_{hit}| > z_{\frac{\alpha}{2}}$ (21)

Pengujian serentak dilakukan untuk mengetahui signifikansi parameter $\beta(u_i, v_i)$ terhadap variabel respon secara bersama-sama pada model GWLR.

H0: $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_P(u_i, v_i) = 0$

H1: paling tidak terdapat satu $\beta_k(u_i, v_i) \neq 0$; $k = 1, 2, \dots, p$

Statistik uji yang digunakan adalah :

$$D(\beta(u_{I,} v_i)) = -2 \ln \left[\frac{L(w)}{L(\pi)} \right] \tag{22}$$

Kriteria penolakan H_0 adalah apabila $D(\beta(u_{I,} v_i)) > x^2{}_{v(a)}$

2.6 Pembobot dan penentuan Bandwidth Optimum

Pemelihan Pembobot spasial yang digunakan dalam menduga parameter sangat penting untuk menentukan besarnya pembobot masing-masing lokasi yang berbeda. Pembobot yang digunakan adalah fungsi kernel sebagai berikut (Fotheringham, et al)

a. Fungsi Gaussian Kernel

$$w_i(u_i v_i) = (1-(d_{ij}/h)^2)^2 \tag{23}$$

b. Fungsi Bisquare Kernel

$$w_i(u_i v_i) = (1-(d_{ij}/h)^2)^2 \qquad jikad_{ij} < h \tag{24}$$

$$jikad_{ij} < h$$

Dengan d_{ij} adalah jarak *euclidean* antara lokasi *I* dan lokasi j dengan menggunakan persamaan:

$d_{ij}\sqrt{(u_i-u_j)^2+(v_i-v_j)^2}$ h adalah parameter penghalus (bandwidth). (25)

Untuk mendapatkan *bandwidth optimum* maka dapat dilakukan dengan menghitung *cross validation* (CV). Jika nilai CV semakin kecil maka akan didapatkan bandwidth yang optimum.

$$cv(h) = \sum_{i=1}^{n}(y_i - y_{\neq i}(h))^2 \tag{26}$$

2.7 Pengujian Kesamaan Model Regresi Logistik dan GWLR

Pengujian parameter digunakan untuk mengetahui parameter yang berpengaruh signifikan terhadap model. Pengujian pertama adalah pengujian kesamaan model model regresi logistik dengan model GWLR untuk mengetahui apakah model GWLR lebih sesuai digunakan dibandingkan dengan model regresi logistik, dengan hipotesis sebagai berikut: yaitu

Ho : $\beta_k(u_i, v_i) = \beta_k : k = 1, 2, \dots, p$ (tidak ada perbedaan yang signifikan antara model regresi logistik dan GWLR)

H1: paling sedikit ada satu $\beta_k(u_i, v_i)$ yang berhubungan dengan lokasi (u_i, v_i) (ada perbedaan yang signifikan antara model logistik dan GWLR).

Statistik ujinya:

$$F\ hitung = \frac{devians\ model\ 1/df1}{devians\ model\ 2/df2} \qquad (27)$$

2.8 Pengujian Asumsi Non Multikolinieritas dan Heterogenitas Spasial

Asumsi non-multikolinieritas mengharuskan tidak adanya korelasi antara satu peubah prediktor dengan peubah prediktor lainnya. Salah satu cara untuk mendeteksi adanya multikolinieritas dalam data adalah dengan melihat nilai Varian Inflation Factor (VIF). Apabila nilai VIF > 10, maka dapat dikatakan bahwa terdapat multikolinieritas. Persamaan dari VIF dituliskan sebagai berikut (Kutner, et. al.,2005):

Heterogenitas Spasial terjadi akibat adanya perbedaan antara satu wilayah dengan wilayah lainnya. Pengujian heterogenitas Spasial menggunakan uji *Breusch-Pagan.* Hipotesisnya sebagai berikut :

$H_0 : \sigma_1^{\ 2} = \sigma_2^{\ 2} \dots = \sigma_n^{\ 2} = \sigma^2$

H_1 : minimal terdapat satu $\sigma_1^{\ 2} \neq \sigma_2^{\ 2}$

statistika uji Breush – pagan

$$BP = \frac{1}{2} f^{i} (Z(Z^{i}Z)^{-1} Z^{i} f + \left(\frac{1}{T}\right)\left(\frac{e^{i}we}{\sigma^{2}}\right)^{2} \quad (28)$$

dengan vektor **f** adalah

$$\boldsymbol{f_i} = \left(\frac{\boldsymbol{ei^1}}{\boldsymbol{\alpha^2}}\right) - \mathbf{1} \quad (29)$$

2.9 Pemilihan Model Terbaik

AIC (*Akaike Information Criterion*) adalah salah satu kriteria pemilihan model terbaik dengan mempertimbangkan banyaknya parameter [3]. Kriteria AIC dapat dituliskan sebagai berikut :

$$AIC = 2n \ln(\sigma) + \ln(2\pi) + n + tr(s) \quad (30)$$

Dengan :

ln : natural log

σ : nilai estimator standar deviasi dari bentuk residual

n : banyaknya pengamatan

π : 3,14

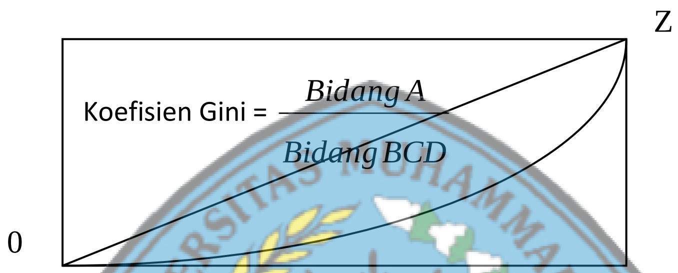
S : matriks proyeksi

Model terbaik adalah model yang memiliki nilai AIC terkecil.

2.10 Rasio Gini

Koefisien gini (Gini Ratio) adalah satu ukuran yang paling sering digunakan untuk mengukur ketidak merataan atau tingkat ketimpangan agregat (

secara keseluruhan) yang angkanya berkisar antara nol (pemerataan sempurna) hingga satu (ketimpangan yang sempurna). Rasio Gini merupakan indkator untuk menilai suatu ketimpangan. Koefisien Gini dapat diperoleh dengan menghitung rasio bidang yang terletak antara garis diagonal dan kurva Lorenz dibagi dengan luas separuh bidang di mana kurva Lorenz itu berada. .



Gambar 2.1 Kurva Lorenz dan Koefisien Gini

Semakin jauh jarak garis kurva *Lorenz* dari garis diagonal, semakin tinggi tingkat ketidakmerataannya. Sebaliknya semakin dekat jarak kurva Lorenz dari garis diagonal, semakin tinggi tingkat pemerataan nya. Pada gambar di atas, besarnya ketimpangan digambarkan sebagai daerah yang diarsir.

Tingkat rasio gini (ketimpangan) menurut (Lincolin Arsyad,1997) dibagi menjadi 3 yaitu :

1. Tingkat ketimpangan rendah 0,20 – 0 ,35
2. Tingkat ketimpangan sedang 0,36 – 0,49
3. Tingkat ketimpangan Tinggi 0,50 – 0,70

Jika nilai gini rasio mendekati 0 artinya ketimpangan semakin kecil. Sebaliknya, apabila mendekati 1 artinya ketimpangan semakin besar.

Faktor-faktor yang mempengaruhi rasio gini

a. Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) / Human Development Index (HDI) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara seluruh dunia. IPM digunakan untuk mengklasifikasikan apakah sebuah negara adalah negara maju, negara berkembang atau negara terbelakang dan juga untuk mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup. UNDP (*United Nation Development Programme*) mendefinisikan pembangunan manusia sebagai suatu proses untuk meperluas pilihan -pilihan bagi penduduk. Semakin cepat pembangunan manusia dengan cara pemerataan pendidikan dan kesehatan maka pertumbuhan ekonomi akan mencapai peningkatan produktivitas dan kesempatan kerja.

b. Pertumbuhan Ekonomi

Pertumbuhan ekonomi adalah perkembangan kegiatan dalam perekonomian yang menyebabkan barang dan jasa yang diproduksikan dalam masyarakat bertambah dan kemakmuran masyarakat meningkat. Pertumbuhan ekonomi berbeda dengan pembangunan ekonomi karena dalam pembangunan ekonomi tingkat pendapatan perkapitanya terusmenerus meningkat sedangkan pertumbuhan ekonomi belum tentu diikuti oleh kenaikan pendapatan perkapita (Sukirno, 2006).

Pertumbuhan ekonomi daerah adalah pertambahan pendapatan masyarakat secara keseluruhan yang terjadi di daerah tersebut, yaitu kenaikan seluruh nilai tambah yang terjadi. Perhitungan pendapatan daerah awalnya dibuat dalam harga berlaku, namun agar dapat melihat pertambahan dari satu kurun waktu ke kurun waktu berikutnya, harus dinyatakan dalam nilai rill, artinya dinyatakan dalam harga konstan (Tarigan, 2005)

c. Tenaga Kerja

Badan Pusat Statistika (BPS) mendefinisikan tenaga kerja sebagai Penduduk usia 15 tahun ke atas yang sedang bekerja, yang memiliki pekerjaan namun sementara tidak bekerja, seseorang yang tidak memiliki pekerjaan dan sedang mencari pekerjaan dikategorikan bekerja. Lewis mengemukakan bahwa ada dua sektor di dalam perekonomian negara sedang berkembang, yaitu sektor modern dan sektor tradisional.Sektor informal mampu menyerap kelebihan tenaga kerja yang ada selama berlangsungnya proses industrialisasi, sehingga disebut katub pengaman ketenagakerjaan. Dengan terserapnya kelebihan tenaga kerja disektor industri (sektor modern) oleh sektor informal, maka pada suatu saat tingkat upah di pedesaan akan meningkat. Peningkatan upah ini akan mengurangi perbedaan tingkat pendapatan antara pedesaan dan perkotaan, sehingga kelebihan penawaran pekerja tidak menimbulkan masalah pada pertumbuhan ekonomi(Nizar dkk, 2013).