ISSN 0215-1952

PENGGUNAAN MODEL LOGIT UNTUK MENDUGA PELUANG TERJADINYA HUJAN (Studi kasus hujan di Banjarbaru-Kalimantan Selatan)

Dede Tarmana
Sub Bidang Bina Operasi Perubahan Iklim
Badan Meteorologi Klimatologi dan Geofisika, Jakarta

Sarasanti
Sub Bidang Pencemaran Udara
Badan Meteorologi Klimatologi dan Geofisika, Jakarta

ABSTRAK

Indonesia merupakan wilayah dengan iklim tropik karena terletak di daerah khatulistiwa, sehingga mempunyai dua musim yaitu musim penghujan dan kemarau. Pendugaan peluang suatu hari akan terjadi hujan atau tidak, informasinya sangat bermanfaat untuk aktifitas makhluk hidup. Model yang dapat digunakan untuk hal tersebut (bersifat kategorik) salah satunya adalah model regresi logistik. Dengan menggunakan analisis regresi logistik diharapkan terbentuk model yang baik dalam melakukan pendugaan peluang terjadinya hujan. Dari hasil analisis regresi logistik, diperoleh model terbaik dari dua kandidat model yang dihasilkan yaitu

$$P(Y=0)=\frac{e^{32.562-0.272X_1-0.396X_2}}{1+e^{32.562-0.272X_1-0.396X_2}}$$

(X_1 = Kelembaban pada t-1, X_2 = Suhu pada t-1) dengan nilai AIC = 1435.631 dan SC = 1451.448.

Kata kunci : hujan, model, musim, pendugaan, regresi logistik

1. PENDAHULUAN

Siklus hidrologi memastikan bahwa komponen hujan merupakan bagian yang tak terpisahkan dalam komponen yang menyusun kehidupan di alam semesta ini, secara alami keseimbangan alam akan selalu terjadi melalui adanya fenomena yang berpasangan misalnya : hujan vs kemarau, daratan vs lautan, siang vs malam, hal ini sudah “given” diciptakan oleh sang pencipta. Kajian yang menarik terhadap alam adalah bagaimana mempelajari segala perilaku alam sehingga dapat menghasilkan suatu informasi yang bermanfaat dalam menunjang kehidupan manusia. Dalam kajian geografis, Indonesia yang terletak di daerah khatulistiwa menjadikannya negara beriklim tropis, yang ditandai dengan curah hujan hampir sepanjang tahun terjadi dan intensitasnya yang cukup tinggi dan. fenomena hujan merupakan fenomena alam yang dampaknya terasa langsung oleh manusia dalam beraktifitas, terjadi dengan diawali oleh adanya penguapan dari berbagai benda didaratan dan lautan kemudian membentuk awan diatmosfer untuk selanjutnya melalui proses-proses fisis turunlah hujan.

Gambaran sederhana proses terjadinya hujan tersebut terwakili oleh data-data unsur meteorologi yang dicatat oleh petugas (observer) secara kontinyu, sehingga dalam mempelajari fenomena hujan bisa dilakukan dengan menganalisa perilaku data meteorologi yang tercatat. Dalam melakukan analisa data banyak alat analisa yang dapat digunakan, dalam hal ini penulis akan menggunakan alat dari bidang keilmuan statistika yaitu pemodelan logistic (Regresi Logistik).

Model regresi logistik adalah model regresi yang peubah terikat/ responnya mensyaratkan berupa peubah kategorik sedangkan menurut ***Hosmer(1989)***, Metode regresi logistik adalah suatu metode analisis statistika yang mendeskripsikan hubungan antara peubah respon yang memiliki dua kategori atau lebih dengan satu atau lebih peubah penjelas berskala kategori atau interval. Yang dimaksud dengan peubah kategorik yaitu peubah yang berupa data nominal dan ordinal.

Pendekatan model persamaan regresi logistik digunakan karena dapat menjelaskan hubungan antara *X* dan $\pi(x)$ yang bersifat tidak linear, ketidaknormalan sebaran dari *Y*, keragaman respon yang tidak konstan dan tidak dapat dijelaskan oleh model regresi linear biasa (***Agresti, 1990***). Untuk kasus suatu hari terjadi hujan atau tidak merupakan kejadian biner, dimana model umum regresi logistiknya adalah :

$$g(\mu_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

2. RUMUSAN DAN BATASAN MASALAH

Batasan masalah dari penelitian ini yaitu bahwa data yang digunakan berupa data meteorologi hasil pengamatan di stasiun klimatologi Banjarbaru – Kalimantan Selatan dari tahun 1990-2000 dengan rumusan masalah sebagai berikut:

- Apakah model regresi logistik untuk menduga peluang suatu hari akan turun hujan atau tidak mempunyai tingkat keakuratan yang baik ?

- Apakah parameter regresi logistik yang dihasilkan signifikan melalui uji individu dan keseluruhan model ?

3. TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah seperti berikut:

- Membuat model regresi logistik yang dapat digunakan untuk menduga peluang apakah suatu hari akan turun hujan atau tidak dengan menggunakan variabel penduga data unsur-unsur meteorologi yang secara teoritis dan faktual mewakili perilaku proses terjadinya hujan.

- Menguji model yang sudah dihasilkan baik secara individu maupun keseluruhan.

4. TINJAUAN

4.1. Regresi Logistik

Pada kasus-kasus penelitian dengan tujuan untuk mengetahui hubungan antara suatu peubah dengan peubah penyebab dimana peubah terikatnya berupa data kategorik, maka analisis regresi linear standar tidak bisa dilakukan, oleh karena itu salah satu pendekatan yang dapat dilakukan adalah regresi logistik.
Model persamaan regresi logistik

digunakan untuk dapat menjelaskan hubungan antara *X* dan $\pi(x)$ yang bersifat tidak linear, ketidaknormalan sebaran dari *Y*, keragaman respon yang tidak konstan dan tidak dapat dijelaskan oleh model regresi linear biasa (***Agresti, 1990***). Metode regresi logistik adalah suatu metode analisis statistika yang mendeskripsikan hubungan antara peubah respon yang memiliki dua kategori atau lebih dengan satu atau lebih peubah penjelas berskala kategori atau interval (***Hosmer dan Lemeshow, 1989***).

Jika data hasil pengamatan memiliki *p* peubah bebas yaitu $x_1, x_2, ..., x_p$ dengan peubah respon *Y*, dengan *Y* mempunyai dua kemungkinan nilai 0 dan 1, *Y* = 1 menyatakan bahwa respon memiliki kriteria yang ditentukan dan sebaliknya *Y* = 0 tidak memiliki kriteria, maka peubah respon *Y* mengikuti sebaran Bernoulli dengan parameter $\pi(x_i)$ sehingga fungsi sebaran peluang :

$$f(y_i) = [\pi(x_i)]^{y_i}[1-\pi(x_i)]^{1-y_i}, \; y_i = 0,1$$

Model umum regresi logistik dengan *p* peubah penjelas yaitu

$$\pi(x) = \frac{\exp(g(x))}{1+\exp(g(x))}$$

dengan melakukan transformasi logit diperoleh

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

dengan

$$g(x) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p \; , g(x)$$

merupakan penduga logit yang berperan sebagai fungsi linear dari peubah penjelas. Karena fungsi penghubung yang digunakan adalah fungsi penghubung logit maka sebaran peluang yang digunakan disebut sebaran logistik (***McCullagh dan Nelder, 1989***).

4.2. Pendugaan Parameter

Ada beberapa metode pendugaan parameter dalam regresi, salah satunya yaitu metode maksimum likelihood. Pendugaan parameter β untuk model regresi logistik biner sederhana dengan satu peubah bebas

$$P(Y=0) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}.$$

bisa menggunakan metode *maximum likelihood*. Secara sederhana dapat disebutkan bahwa metode ini berusaha mencari nilai koefisien yang memaksimumkan fungsi likelihood. Dengan nilai Y yang bersifat biner, kita dapat menggunakan Bernoulli sebagai sebaran variabel Y sehingga fungsi *likelihood* akan berbentuk

$$L = f(y_i) = \prod_{i=1}^{n} [p_i]^{y_i}[1-p_i]^{1-y_i}$$

Dengan :

$$p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}$$

Melalui transformasi logaritma maka operasi perkalian berubah menjadi penjumlahan, kemudian fungsi *likelihood* diganti dengan fungsi *log-likelihood*. Perlu diingat bahwa fungsi logaritma besifat monoton naik, sehingga jika *log-likelihood* mencapai maksimum maka fungsi *likelihood* juga demikian. Bentuk fungsi yang dimaksimumkan adalah

$$LL = Log(L) \ = \sum_{i=1}^{n} \log\left[p_i\right]^{y_i}\left[1-p_i\right]^{1-y_i}$$

$$= \sum_{i=1}^{n} y_i \log\left[p_i\right] + [1-y_i]\log\left[1-p_i\right]$$

Penduga bagi koefisien β diperoleh sebagai solusi bagi permasalahan memaksimumkan LL yang dapat diselesaikan melalui prosedur iterasi bobot kuadrat terkecil (Iteratively Weighted Least Squares = IWLS).

4.3. Ukuran Kebaikan Model

Untuk mengukur tentang kesesuaian model regresi Poisson, ada beberapa ukuran statistik yang dapat dijadikan kriteria dalam penentuan kebaikan model, di antaranya yaitu Pearson Chi-square, Deviance, Uji Rasio likelihood, dan uji lainnya (AIC, BIC).

4.3.1 Pearson Chi-squares

Nilai Pearson chi-squares dapat dihitung melalui persamaan berikut:

$$\sum_{i} \frac{(y_i - \mu_i)^2}{Var\ (Y_i)}$$

untuk dibandingkan dengan nilai X^2_{df} dimana df (degree of freedom) = n - p.

4.3.2 Deviance

Deviance merupakan ukuran lain untuk mengukur kebaikan model regresi logistic biner, nilai deviance dapat dihitung melalui $2(\ell(y;y)-\ell(\mu;y))$, dimana $\ell(\mu;y)$ dan $\ell(y;y)$ adalah log likelihood yang dievaluasi terhadap μ dan y.

Untuk model yang baik/sesuai, deviance juga mempunyai kedekatan dengan distribusi chi-squares dengan derajat bebas n – p. Kondisi lain jika nilai antara Pearson Chi-square dan deviance relative sama dengan derajat bebas n – p, maka model yang dihasilkan kemungkinan mempunyai tingkat kesesuaian yang cukup.

4.3.3 Uji Rasio Likelihood

Keuntungan menggunakan metode maksimum likelihood adalah bahwa uji rasio likelihood dapat di implementasikan untuk menaksir kesesuaian dari kelebihan pendugaan parameter regresi logistik dengan menggunakan MLE (Maksimum Likelihood Estimation).

Formula uji rasio likelihood adalah $G=2(\ell_1-\ell_0)$ dimana ℓ_1=likelihood tanpa peubah bebas dan ℓ_0=likelihood dengan peubah bebas. Nilai G mempunyai kedekatan dengan distribusi chi-squrae berderajat bebas k ($G \approx X^2_k$) dengan hipotesis :

H$_0$: $\beta_1 = \beta_2 = ... = \beta_k = 0$
H$_1$: Min ada satu $\beta \neq 0$

4.3.4 Uji lainnya

Ukuran lain yang dapat mengukur kebaikan model adalah AIC dan BSC. AIC (Akaike information criteria) didefinisikan sebagai , dimana ℓ merupakan log likelihood yang dievaluasi terhadap $\mu\ dan\ p$ (p = jumlah parameter), Nilai AIC yang semakin kecil mengindikasikan bahwa model yang baik.

Selain AIC ada juga ukuran kebaikan model lainnya yaitu BSC (Bayesian-Schwartz criteria) yang didefinisikan oleh persamaan berikut:

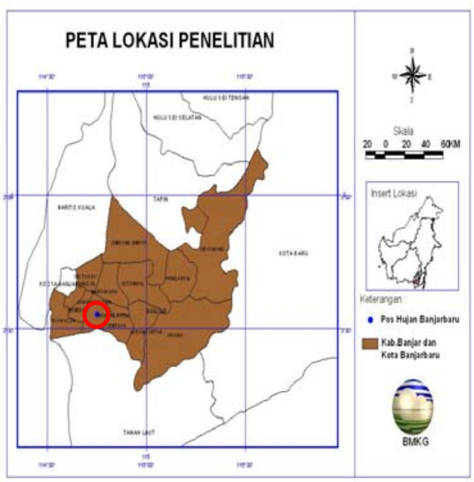
$$BSC = \ell - p \log \left(\frac{n}{2\pi} \right)$$

dimana ℓ merupakan log likelihood yang dievaluasi terhadap μ, p (p = jumlah parameter) dan n (jumlah klas). Kebalikan dari nilai AIC, bahwa apabila nilai BSC semakin besar mengindikasikan bahwa model yang baik.

5. DATA

5.1. Lokasi Penelitian

Secara geografis stasiun klimatologi Banjarbaru – Kalimantan Selatan yang menjadi lokasi penelitian terletak pada 3.45 LS 114.75 BT dengan ketinggian ±12 m DPL seperti tampak pada gambar 1 di bawah ini :



Gambar 1. Peta lokasi penelitian

5.2. Data Penelitian

Data yang digunakan yaitu data Curah Hujan harian, temperatur minimum harian, dan kelembaban harian Stasiun Klimatologi Klas Banjarbaru – Kalimantan Selatan tahun 1997 s/d 2000.

Data curah hujan harian dikategorikan menjadi dua kategori yaitu hujan (0) dan tidak hujan (1) sehingga peubah curah hujan dengan tanpa memandang intesitas hujannya berubah menjadi peubah data biner $\begin{cases} 0 = hujan \\ 1 = tidakhujan \end{cases}$

6. HASIL DAN PEMBAHASAN

Dengan menggunakan pendekatan model regresi logistik diperoleh beberapa kandidat model pendugaan peluang hujan sebagai berikut :

6.1 Model 1

$$P(Y=0) = \frac{e^{31.839 - 0.263X_1 - 0.391X_2 - 0.082X_3 - 0.142X_4 - 0.174X_5}}{1 + e^{31.839 - 0.263X_1 - 0.391X_2 - 0.082X_3 - 0.142X_4 - 0.174X_5}}$$

Di mana :

P(Y=0)	=	Peluang suatu hari tidak hujan
X_1	=	Kelembaban pada t-1
X_2	=	Suhu pada t-1
X_3	=	CH pada t-1
X_4	=	CH pada t-2
X_5	=	CH pada t-3

Model.1 ini memasukan lima peubah bebas yang mungkin berpengaruh terhadap suatu hari akan turun hujan sehingga modelnya terlihat kompleks/tidak sederhana. Interpretasi pengaruh masing-masing peubah bebas terhadap peluang kejadian peubah responnya adalah misal untuk peubah kelembaban : bila kelembaban naik satu satuan dan peubah lain tetap maka peluang Y=0 (suatu hari tidak hujan) meningkat sebanding dengan exponential dari kelembaban, begitupun interpretasi dari peubah bebas lainnya.

6.2 Model 2

$$P(Y=0)=\frac{e^{32.562-0.272X_1-0.396X_2}}{1+e^{32.562-0.272X_1-0.396X_2}}$$

dimana :

P(Y=0)	=	Peluang suatu hari tidak hujan
X_1	=	Kelembaban pada t-1
X_2	=	Suhu pada t-1

Model.2 hanya melibatkan dua peubah bebas yang secara teoritik mempunyai pengaruh terhadap peluang terjadinya hujan yaitu kelembaban dan temperatur minimum. Interpretasi dari model.2 serupa dengan model.1 yaitu bila kelembaban naik satu satuan dan temperatur minimum tetap maka peluang Y=0 (suatu hari tidak hujan) meningkat sebanding dengan exponential dari kelembaban, interpretasi yang sama untuk untuk peubah temperatur.

6.3 Pemilihan Model Terbaik

Dari kedua kandidat model di atas, maka dapat diambil satu model terbaik yang dapat dipergunakan untuk menghitung peluang dugaan suatu hari akan turun hujan atau tidak. Seperti telah dikatakan pada tinjauan pustaka bahwa ada beberapa uji yang bisa dilakukan untuk mendapatkan suatu model dari beberapa kandidat model di antaranya Uji kelayakan model secara menyeluruh (goodness of fit), Uji parameter secara individu, atau dengan melihat nilai deviance, AIC, BIC dari masing-masing model yang diperbandingkan.

Tabel 6.1 Uji Kebaikan model

Model	G (Likelihood Ratio)	P_value
1	565.1659	<0.0001
2	562.4055	<0.0001

Dari hasil pengujian secara menyeluruh, tampak bahwa kedua model mempunyai kelayakan dalam menduga peluang terjadinya hujan atau tidak, namun demikian masih perlu diuji parameter secara individu dari masing-masing model.

Tabel 6.2 Uji Parameter secara Individu

Paremeter	Chi Square	P_value
Model.1		
RH_{t-1}	226.6854	0.0001
$Tmin_{t-1}$	42.3359	0.0001
CH_{t-1}	0.3738	0.5410
CH_{t-2}	1.1328	0.2872
CH_{t-3}	1.7150	0.1903
Model.2		
RH_{t-1}	262.9947	0.0001
$Tmin_{t-1}$	44.0945	0.0001

Dari hasil uji parameter secara individu masing-masing model pada α = 5%, tampak bahwa untuk model.1 terdapat parameter yang tidak signifikan yaitu CH_{t-1},CH_{t-2} dan $CH_{t-3,}$ sedangkan untuk model2 semua parameternya signifikan, sehingga berdasarkan uji ini model.2 lebih baik dari model.1.

Tabel 6.3 Ukuran Fitting Model

Model	AIC	SC
1	1435.872	1467.499
2	1435.631	1451.448

Dari informasi nilai AIC dan SC, model 2 mempunyai nilai yang lebih kecil dibanding dengan model1, berdasarkan hal ini berarti bahwa model 2 mempunyai kelayakan lebih baik daripada model 1.

7. KESIMPULAN DAN SARAN

7.1. Kesimpulan

- Secara Umum kelembaban dan temperatur minimum cukup berpengaruh terhadap dugaan peluang turun atau tidaknya hujan pada suatu hari, walaupun kemungkinan ada variabel-variabel lain berpengaruh tetapi tidak masuk dalam model.

- Terjadinya hujan pada suatu hari tidak terlepas dari rangkaian kejadian hujan sebelumnya (CH_{t-1},CH_{t-2} dan CH_{t-3}), hal ini menjadi dasar peubah curah hujan hari sebelumnya menjadi regresor /penduga dalam model.

- Dari dua kandidat model, berdasarkan uji-uji statistik dan ukuran kebaikan model maka **model.2** lebih baik dibanding dengan **model.1**.

7.2. Saran

- Untuk lebih meningkatkan kualitas model, perlu dilakukan analisa lebih lanjut mengenai peubah - peubah yang berpengaruh terhadap respon (hujan) selain peubah yang sudah ada pada model.

- Model regresi logistik biner hasilnya hanya berupa dugaan hujan atau tidak, tanpa menduga intensitas hujan yang akan terjadi, sehingga perlu pengembangan dari biner ke multinomial.

8. ACUAN

Agresti, Allan. 1990. ***Categorical Data Analysis***. New York: John Wiley and Sons.

Agresti, Allan. 1996. ***An Introduction to Categorical Data Analysis***. New York: John Wiley and Sons

Bret Larget. 2007. "***Poisson Regresion***". J. Dept of Botani & Statistics. Universitas of Wisconsin. Madison

.Hosmer, D.W. dan S. Lemeshow. 1989. **Applied Logistic Regression**. New York: John Wiley and Sons.

McCullagh, P. dan Nelder, J. A. 1989. ***Generalized Linear Models 2nd Edition***. London: Chapman & Hall.