

BINARY LOGISTIC REGRESSION (BLR) TERHADAP STATUS BEKERJA DI KOTA SURABAYA

Moh. Yamin Darsyah¹ Arianto Wijaya²

^{1,2}Program Studi S1 Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

*Email mydarsyah@unimus.ac.id

Abstract

One of method categorical Data Analysis used to know the response is used to determine the effect the relationship with the response variable of type face is Binary Logistic Regression (BLR). BLR used for the response variable data is data that consists of two categories, with one or predictor variables more, both categorical and kontinu. The research aims determine the degree of labor in Indonesia, one of the indicators that can be used is the status defined working. Status working into two, namely the status is still working and status not working. Variabel used in this study is the status working (Y) as the response variable with category 1 is still working, category 0 not working, whereas the predictor variable is the level of education (X1) and gender (X2). The result is There are two variables influencing the model of the variable educational level (X1) and gender (X2).

Keywords: BLR, Labor, Status Working

1. PENDAHULUAN

Analisis regresi adalah suatu metode statistika yang umum digunakan untuk melihat pengaruh antara peubah bebas (variabel prediktor) dengan peubah tak bebas (variabel respon). BLR digunakan untuk menjelaskan hubungan antara variabel respon yang dikotomi/biner dengan variabel bebas yang berupa data berskala interval dan atau kategorik (Holmes dan Lemeshow, 1989). Variabel yang dikotomi/biner adalah variabel yang hanya mempunyai dua kategori saja, yaitu kategori yang menyatakan kejadian sukses (Y=1) dan kategori yang menyatakan gagal (Y=0). BLR dapat digunakan untuk mengukur tingkat akurasi dari klasifikasi data dalam hal ini data jenis dikotomi/biner sehingga hasil akurasi nya variabel respon mendekati akurat. Beberapa penelitian yang pernah dilakukan untuk kasus data biner antara lain Darsyah (2013) menakar tingkat akurasi *support vector machine* untuk kasus kanker payudara, Palupi dan Abadyo (2013) Perbandingan Regresi Model Logistik Biner dengan Regresi Model Probit terhadap Faktor – Faktor yang Mempengaruhi Sikap Siswa SMP pada Pelajaran Matematika.

Pembangunan sumber daya manusia menjadi salah satu fokus pemerintah dalam memerangi kemiskinan serta pemetaan tenaga kerja. Tenaga kerja yang berkualitas memberi dampak pada kesejahteraan masyarakat. Dalam menentukan derajad tenaga kerja di Indonesia, salah satu indikator yang dapat digunakan adalah status bekerja. Status bekerja didefinisikan menjadi dua, yaitu status masih bekerja dan status tidak bekerja (BPS). Adapun tujuan dari penulisan ini adalah untuk mengetahui pengaruh tingkat pendidikan dan jenis kelamin terhadap status bekerja pada metode regresi logistik biner di Kota Surabaya.

2. KAJIAN LITERATUR

1.1 Analisis Regresi

Analisis regresi adalah suatu metode statistika yang umum digunakan untuk melihat pengaruh antara peubah bebas (variabel prediktor) dengan peubah tak bebas (variabel respon). Misa lnya Y adalah variabel respon dan X adalah prediktor, secara umum hubungan antara Y dan X dapat ditulis sebagai berikut :

$$Y_i = f(X_i) + \epsilon_i, \quad i = 1, 2, 3, \dots, n \quad (2.1)$$

1.2 Regresi Logistik

Model regresi logistik yang dipengaruhi oleh p variabel prediktor dapat dinyatakan sebagai nilai harapan dari Y dengan diberikan nilai x

$$E(Y|x) = \frac{exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (2.2)$$

Dengan $0 \leq E(Y|x) \leq 1$ dan Y mempunyai nilai 0 atau 1. Nilai 1 merupakan probabilitas $E(Y|x)$ sukses, sehingga dapat dinyatakan dengan $\pi(x)$, sehingga persamaan diatas menjadi

$$\pi(x) = \frac{exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (2.3)$$

Dengan β_k menyatakan parameter – parameter regresi x_k adalah pengamatan variabel prediktor (Hosmer dan Lemeshow (1989) dalam Tiro (2000)) Transformasi logit diterapkan pada model regresi logistik,

$$Logit(\pi(x)) = g(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \sum_{k=1}^p \beta_k x_k \quad (2.4)$$

Keterangan :

$\pi(x)$ adalah peluang kejadian.

$g(x)$ adalah nilai estimasi logit

$\beta_0, \beta_1, \dots, \beta_p$ berturut-turut adalah nilai koefisien untuk variabel – variabel konstan, tingkat pendidikan dan jenis kelamin yang diperoleh menggunakan metode *Maximum Likelihood Estimation*.

Transformasi logit bertujuan untuk membuat fungsi linier dari parameter – parameter nya. Fungsi $g(x)$ linier terhadap parameter dan memiliki range $(-\infty, \infty)$, tergantung dari range variabel prediktor X .

1.3 Binary Logistic Regression (BLR)

BLR adalah suatu regresi logistik antara variabel respon (y) dan variabel prediktor (x) dimana variabel y menghasilkan 2 kategori yaitu 0 dan 1 (Hosmer dan Lemeshow, 1989). Sehingga variabel y mengikuti distribusi

Bernoulli dengan fungsi probabilitasnya sebagai berikut.

$$f(y) = \pi^y (1 - \pi)^{1-y}; y = 0, 1 \quad (2.5)$$

Dimana jika $y = 0$ maka $f(y) = 1 - \pi$ dan jika $y = 1$ maka $f(y) = \pi$. Fungsi regresi logistiknya dapat dituliskan sebagai berikut.

$$f(z) = \frac{1}{1+e^{-z}} \text{ ekuivalen } f(z) = \frac{e^z}{1+e^z} \quad (2.6)$$

Dengar $z = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$

Jika nilai z antara $-\infty$ dan ∞ maka nilai $f(z)$ terletak antara 0 dan 1 untuk setiap nilai z yang diberikan. Hal tersebut menunjukkan bahwa model logistik sebenarnya menggambarkan probabilitas atau resiko dari suatu obyek. Model regresi logistiknya adalah sebagai berikut.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (2.7)$$

Dimana p = banyaknya variabel prediktor

Bila model persamaan di atas ditransformasi dengan transformasi logit, maka didapatkan bentuk logit seperti pada persamaan (2.8).

$$g(x) = \ln \left(\frac{\pi(x)}{1-\pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.8)$$

1.4 Status Bekerja

Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tanpa jeda) dalam sehari (BPS).

1.5 Tingkat Pendidikan

Menurut Palupi & Abadyo (2013) tingkat pendidikan adalah suatu proses jangka panjang yang menggunakan prosedur sistematis dan terorganisir.

Tingkat pendidikan memiliki tahapan pendidikan yang ditetapkan berdasarkan tingkat perkembangan peserta didik, tujuan yang akan dicapai dan kemauan yang dikembangkan (Suharjo, 2007) terdiri dari :

Pendidikan dasar adalah jenjang pendidikan awal selama 9 (sembilan) tahun pertama masa sekolah anak yang melandasi jenjang pendidikan menengah, meliputi : SD, MI, SMP/SMPLB, MTs.

Pendidikan menengah adalah jenjang lanjutan pendidikan dasar, meliputi : SMA/SMALB, MA, SMK.

Pendidikan tinggi merupakan jenjang pendidikan setelah pendidikan menengah yang mencakup program sarjana, magister, doktor, dan spesialis yang diselenggarakan oleh perguruan tinggi.

1.6 Jenis Kelamin

Jenis Kelamin (*sex*) adalah perbedaan antara perempuan dan laki – laki secara biologis sejak seseorang lahir. Jenis kelamin berkaitan dengan tubuh laki – laki dan perempuan, dimana tubuh laki – laki memproduksi sperma, sementara perempuan menghasilkan sel telur dan secara biologis mampu untuk menstruasi, hamil, dan menyusui. Perbedaan biologis dan fungsi biologis laki – laki dan perempuan tidak dapat dipertukarkan diantara keduanya.

3. METODE PENELITIAN

a. Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini adalah data sekunder dari hasil Data Susenas Kota Surabaya tahun 2012. Dalam penelitian ini ada sebanyak 3303 responden. Variabel yang digunakan dalam penelitian ini adalah status bekerja (Y) sebagai variabel respon dengan kategori 1 adalah masih bekerja, kategori 0 tidak bekerja, sedangkan variabel prediktornya yaitu tingkat pendidikan (X1) dan jenis kelamin (X2).

b. Metode Analisis

Langkah – langkah analisis yang dilakukan pada penelitian ini adalah :

Pengambilan data dan Melakukan analisis BLR dengan software SPSS 21.0 untuk pembentukan

model BLR. Langkah analisisnya sebagai berikut :

- Pembentukan model BLR
- Mengasumsikan variabel respon Y berdistribusi bernoulli dalam BLR.
- Memeriksa kejadian pemisahan data pada BLR.
- Melakukan uji signifikansi parameter beta dengan menggunakan uji *maximum likelihood ratio test* dan dilanjutkan dengan uji *Wald*.
- Melakukan uji kecocokan *Hosmer and Lemeshow Goodness of Fit Test*.
- Melakukan interpretasi dalam pengukuran asosiasi & sensitifitas.
- Menginterpretasi model status bekerja dan variabel – variabel yang berpengaruh didalam model.

4. HASIL DAN PEMBAHASAN

4.1 Estimasi Parameter

Diperoleh estimasi parameter regresi logistik, sehingga model BLR dapat dituliskan sebagai berikut:

$$\text{logit}(P_i) = -0.997 - 0.208 \text{ tingkat pendidikan} +$$

dimana:

$$P_i = P(Y = 1)$$

$$g(x) = \text{logit}(P) = \log\left(\frac{P}{1-P}\right)$$

sehingga

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

4.2 Uji Signifikansi Parameter

a. Uji serentak parameter regresi logistik
Adapun hipotesis untuk pengujian signifikansi parameter regresi secara serentak yaitu:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_5 = 0$$

$$H_a : \text{Minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, n$$

Statistik uji yang digunakan yaitu uji χ^2 . H_0 ditolak bila $p\text{-value} < \alpha$, untuk $\alpha = 0.05$. Dari Tabel 5, $\chi^2 = 524.536$ dan $p\text{-value} = 0.000$, sehingga dapat disimpulkan untuk menolak H_0 . Jadi minimal ada satu parameter regresi logistik

tidak sama dengan nol. Tabel 1 diketahui nilai sig pada uji omnibus menunjukkan ada pengaruh signifikan secara serentak variabel predictor terhadap variabel respon.

Tabel 1. Uji Omnibus Koefisien Model

| | <i>Chi-square</i> | <i>df</i> | <i>Sig.</i> |
|-------|-------------------|-----------|-------------|
| Step | 524.536 | 2 | .000 |
| Block | 524.536 | 2 | .000 |
| Model | 524.536 | 2 | .000 |

b. Uji Parsial Parameter Regresi Logistik

Adapun hipotesis untuk pengujian signifikansi parameter regresi secara parsial yaitu:

$$H_0 : \beta_j = 0$$

$$H_a : \beta_j \neq 0; j = 1, 2, \dots, 5$$

Statistik uji yang digunakan yaitu uji *chi-square*, H_0 ditolak apabila $\chi^2_{hitung} > \chi^2_{Tabel(\alpha/2)}$ atau jika $p\text{-value} < \alpha$, untuk $\alpha = 0.05$. Disimpulkan bahwa secara uji parsial terhadap pengaruh variabel tingkat pendidikan dan jenis kelamin terhadap status bekerja.

4.3 Uji Kesesuaian Model

Tahap selanjutnya yaitu menguji kesesuaian model (*goodness of fit*). Adapun hipotesis dari uji kesesuaian model yaitu:

H_0 : model yang dihipotesakan sesuai dengan data

H_a : model yang dihipotesakan tidak sesuai dengan data

Tabel 2 menunjukkan pengujian kesesuaian model BLR. Untuk pengujian ditampilkan uji *Hosmer-Lemeshow*. Dimana $p\text{-value} = 0.000$, lebih kecil bila dibandingkan dengan α ($\alpha = 0.05$), sehingga dapat disimpulkan bahwa model yang dihipotesakan mendekati sesuai dengan data.

Tabel 2. Uji Hosmer dan Lemeshow

| <i>Step</i> | <i>Chi-square</i> | <i>df</i> | <i>Sig.</i> |
|-------------|-------------------|-----------|-------------|
| 1 | 47.595 | 8 | .320 |

4.4 Ekspektasi dan Pengukuran Asosiasi

Pada Tabel 3 dapat dilihat frekuensi amatan dan harapan dari data, dapat dilihat sejauh mana keragaman variabel respon Y dapat dijelaskan oleh variabel prediktor X_i dengan melihat *Nagelkerke R-square*. Pada kasus ini diperoleh nilainya sebesar 49.7% yang berarti bahwa sebesar 49.7% keragaman variabel respon status bekerja dapat dijelaskan oleh variabel prediktor (tingkat pendidikan dan jenis kelamin). Sedangkan sisanya 50,03% dipengaruhi oleh faktor-faktor diluar variabel prediktor.

Tabel 3. Ringkasan Model

| <i>Step</i> | <i>-2 Log Likelihood</i> | <i>Cox & Snell R Square</i> | <i>Nagelkerke R Square</i> |
|-------------|--------------------------|---------------------------------|----------------------------|
| 1 | 4006.078 ^a | .147 | .497 |

4.5 Sensitivitas atau Spesifisitas

Model regresi logistik yang terbentuk bisa membuat klasifikasi dalam penaksiran nilai Y yaitu sebesar 66%. Artinya dengan model persamaan regresi logistik ini bisa memprediksi seseorang masih bekerja dimana pada kenyataannya seseorang memang masih bekerja, atau memprediksi seseorang tidak bekerja dimana pada kenyataannya seseorang memang tidak bekerja adalah sebesar 66.6%. Dalam dunia sosial disebut dengan sensitivitas atau spesifisitas.

5. SIMPULAN

Kesimpulan dari penelitian adalah BLR digunakan untuk data yang variabel responnya merupakan data yang terdiri dari dua kategori, dengan satu variabel prediktor atau lebih, baik yang bersifat kategorik maupun kontinu. Dari kasus status bekerja masyarakat di Kota Surabaya, dapat diambil kesimpulan bahwa status bekerja dapat dipengaruhi oleh variabel tingkat pendidikan dan jenis kelamin. Pengukuran asosiasi pada model sebesar 49,7% dengan tingkat akurasi klasifikasinya sebesar 66,6%.

6. REFERENSI

- Agresti, A. (1996). An Introduction to Categorical Data Analysis, John Wiley and Sons, New York.*
- Casella, G. and Berger, R.L. (2002), Statistik Inference, Duxbury Thomson Learning, USA.*
- Darsyah. M.Y. (2013). Menakar Tingkat Akurasi Support Vector Machine Pada Kasus Kanker Payudara. Jurnal Statistika Vol 1 Nomer 1 2013, Universitas Muhammadiyah Semarang, Semarang.*
- Hosmer, D.W. dan Lemeshow, S. (1989). Applied Logistic Regression, John Wiley & Sons, Inc., New York.*
- Palupi, W.G.G. dan Abadyo. (2013). Perbandingan Regresi Model Logistik Biner dengan Regresi Model Probit terhadap Faktor – Faktor yang Mempengaruhi Sikap Siswa SMP pada Pelajaran Matematika, Universitas Negeri Malang, Malang.*