

Perbandingan Ketepatan Klasifikasi Analisis Regresi Model Logit Biner dan Model Probit Biner terhadap Faktor-Faktor yang Mempengaruhi Risiko Penyakit Jantung Koroner

Fitri Apriliyanti*, Teti Sofia Yanti

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*fitriapriyanti80@gmail.com, tetisofiyanti@gmail.com

Abstract. The method to determine the relationship between the response variable and several predictor variables is linear regression. When the response variable is binary, the method used is binary logistic regression. Several approaches to developing models for explaining binary regression models are logit models and probit models. Logit models and probit models are commonly used to analyze the relationship between one response variable which is categorical data and several predictor variables. In this study, we want to know the best model between the binary logit model and the binary probit model using coronary heart disease risk data in the City of Framingham, America. From the results of the analysis that has been done, it can be concluded that the factors that influence the risk of coronary heart disease are gender, age, active smokers, cholesterol levels, BMI and glucose. The classification accuracy value of the binary logit regression method is 84.95% and the classification accuracy value of the binary probit method is 84.92%. From the comparison of the classification accuracy between the binary logit regression method and the binary probit method are relatively the same.

Keywords: Binary Logit Model Regression, Binary Probit Model Regression, Coronary Heart Disease.

Abstrak. Metode untuk mengetahui hubungan variabel respon dan beberapa variabel prediktor adalah regresi linier. Ketika variabel respon adalah biner, metode yang digunakan adalah regresi logistik biner. Beberapa pendekatan untuk mengembangkan model untuk menjelaskan model regresi biner adalah model logit dan model probit. Model logit dan model probit biasa digunakan untuk menganalisis hubungan antara satu variabel respon yang merupakan data kategorik dan beberapa variabel prediktor. Penyakit jantung kronis merupakan salah satu penyebab kematian nomor satu di dunia. Dalam penelitian ini, ingin mengetahui model terbaik antara model logit biner dan model probit biner dengan menggunakan data risiko penyakit jantung koroner di Kota Framingham Amerika. Dari hasil analisis yang telah dilakukan, dapat disimpulkan bahwa faktor yang mempengaruhi risiko penyakit jantung koroner adalah jenis kelamin, usia, perokok aktif, kadar kolesterol, IMT dan glukosa. Nilai ketepatan klasifikasi metode regresi logit biner sebesar 84,95% dan nilai ketepatan klasifikasi metode probit biner sebesar 84,92%. Dari hasil perbandingan ketepatan klasifikasi antara metode regresi logit biner dan metode probit biner relatif sama.

Kata Kunci: Regresi Model Logit Biner, Regresi Model Probit Biner, Penyakit Jantung Koroner.

1. Pendahuluan

Kesehatan merupakan perspektif penting pada kehidupan manusia, banyak permasalahan terjadi yang menyebabkan kesehatan dapat terganggu dan mengakibatkan munculnya berbagai macam penyakit, diantaranya perubahan gaya hidup masyarakat dunia yang kurang sehat. Salah satu penyakit yang merupakan penyebab kematian nomor satu di dunia yaitu Penyakit Jantung Koroner. Menurut *World Health Organization* (WHO) pada tahun 2005 paling sedikit 17,5 juta kematian di seluruh dunia disebabkan oleh penyakit jantung dan 60% dari seluruh penyebab kematian penyakit jantung adalah Penyakit Jantung Koroner (PJK).

Dalam bidang statistika, metode untuk mengetahui hubungan variabel respon dan beberapa variabel prediktor adalah regresi linier. Syarat untuk memenuhi regresi linier adalah variabel respon yang harus bersifat kuantitatif dengan skala pengukuran minimal interval. Apabila variabel respon bersifat dikotomis (biner) atau polikotomis yang terdiri lebih dari dua kategori dengan sekumpulan variabel prediktor bersifat kategori atau kontinu adalah model regresi logistik (Hosmer & Lemeshow, 1989). Dikarenakan persamaan regresi linier tidak dapat memenuhi persyaratan dengan dua kategori atau biner, maka metode yang digunakan adalah regresi logistik biner. Beberapa pendekatan untuk mengembangkan model yang menjelaskan model regresi biner adalah model logit dan model probit.

Berdasarkan uraian di atas, akan dilakukan penelitian dengan menggunakan variabel respon yang bersifat kategorik biner dan beberapa variabel prediktor yang diduga mempengaruhi variabel respon. Tujuan dilakukannya penelitian ini adalah untuk mengetahui model terbaik antara metode regresi model logit dan model probit. Penerapannya menggunakan data risiko Penyakit Jantung Koroner di Kota Framingham Amerika beserta faktor-faktor yang mempengaruhinya.

2. Landasan Teori

Regresi Logit Biner

Menurut Hosmer dan Lemeshow (2000), regresi logistik merupakan metode analisis data yang menghubungkan antara variabel respon yang bersifat kategorik dengan variabel prediktor yang bersifat kategorik atau kontinu. Regresi logistik dengan variabel respon dari dua kategori disebut dengan regresi logistik biner. Pada regresi logistik biner, data variabel respon terdiri dari dua kategori, misalnya 1 jika sukses dan 0 jika gagal. Sehingga variabel respon mengikuti distribusi bernoulli dengan peluang $P(Y = 1) = \pi$ dan $P(Y = 0) = 1 - \pi$. Maka fungsi peluangnya adalah:

$$\begin{aligned} f(y_i, \pi_i) &= \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\ &= (1 - \pi_i) \left[\frac{\pi_i}{(1 - \pi_i)} \right]^{y_i} \quad ; y_i = 0,1 \end{aligned}$$

Model regresi logistik yang digunakan adalah:

$$\pi(x) = \frac{\exp(\theta_0 + \theta_1 x_1 + \dots + \theta_p x_p)}{1 + \exp(\theta_0 + \theta_1 x_1 + \dots + \theta_p x_p)}$$

Dimana:

$\pi(x)$ = peluang nilai sukses

p = banyaknya variabel prediktor.

Persamaan $\pi(x)$ dapat ditransformasikan menggunakan transformasi logit yang merupakan fungsi hubungan dari model regresi logistik, sehingga menjadi:

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \theta_0 + \theta_1 x_1 + \dots + \theta_p x_p$$

Menurut Agresti (2002) untuk menentukan estimasi parameter pada regresi logistik biner yang membutuhkan turunan pertama dan turunan kedua dari fungsi *likelihood* yaitu *Maximum Likelihood*.

Regresi Probit Biner

Menurut Gurajati (2004), Regresi probit biner merupakan suatu metode regresi yang dapat

digunakan untuk menganalisis variabel respon yang bersifat kategori biner dengan variabel prediktor yang berupa data kontinu atau diskrit berskala nominal atau biner dengan pendekatan *Cummulative Distribution Funcion* (CDF) normal untuk mengestimasi model probit. Menurut Greene (2008), pemodelan dengan menggunakan metode regresi probit biner berasal dari variabel respon kualitatif Y yang dapat dimulai dengan persamaan sebagai berikut:

$$Y = \theta^T x + \varepsilon$$

Dimana:

Y = variabel respon diskrit

θ = vektor parameter koefisien dengan $\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_p]^T$

p = banyaknya variabel prediktor

x = vektor variabel dengan $x = [1 \ X_1 \ \dots \ X_p]^T$

ε = vektor *error* yang diasumsikan berdistribusi $N(0,1)$.

Estimasi parameter dalam persamaan regresi probit biner dapat menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini menaksir parameter dengan memaksimalkan fungsi likelihood dengan syarat data mengikuti distribusi tertentu (Gujarati, 2004).

Pengujian Parameter

1. Uji Serentak

Uji serentak yang digunakan adalah Uji G atau Uji *Ratio Likelihood* dengan pengujian:

Hipotesis:

$H_0 : \theta_1 = \theta_2 = \dots = \theta_p = 0$

H_1 : paling sedikit ada satu $\theta_j \neq 0$

Statistik Uji :

$$G = -2 \ln \left[\frac{L(\theta_0)}{L(\theta)} \right]$$

dimana:

$L(\theta_0)$ = fungsi *likelihood* di bawah H_0 , persamaannya adalah

$L(\theta_0) = \prod_{i=1}^n \{ [1 - \Phi(y - \theta_0)]^{y_i} [\Phi(y - \theta_0)]^{1-y_i} \}$

$L(\theta)$ = fungsi *likelihood* di bawah populasi, yaitu persamaan (2.14)

Nilai Kritis :

$X_{\alpha,p}^2$: nilai kritis diperoleh dari distribusi Chi Kuadrat dengan derajat bebas p , dimana p adalah banyaknya variabel prediktor.

Kriteria Uji :

Kriteria mengambil taraf nyata α maka H_0 ditolak jika $G > X_{\alpha,p}^2$ atau $pvalue < \alpha$.

2. Uji Parsial

Uji parsial yang digunakan adalah Uji Wald dengan pengujian:

Hipotesis:

$H_0 : \theta_j = 0$

$H_1 : \theta_j \neq 0$ dengan $j = 1, 2, \dots, p$

Statistik Uji :

$$W_j = \left\{ \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \right\}^2$$

Dimana:

$\hat{\theta}_j$ = penaksir θ_j dan $SE(\hat{\theta}_j)$ adalah penaksir galat baku θ_j .

Nilai Kritis :

$X_{(\alpha,p)}^2$: pengujian diperoleh dari distribusi Chi Kuadrat dengan derajat bebas p , dimana p adalah banyaknya variabel penjelas.

Kriteria Uji :

Kriteria pengujian mengambil taraf nyata α maka H_0 ditolak jika $W_j > X_{(\alpha,1)}^2$.

3. Uji Kesesuaian Model (*Goodness of Fit*)

Uji Kesesuaian model yang digunakan adalah Uji Hosmer dan Lemeshow dengan pengujian:

Hipotesis :

H_0 : Model sesuai
 H_1 : Model tidak sesuai
 Statistik Uji :

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Dimana:

- O_k : $\sum_{j=1}^{c_k} y_j$ yaitu observasi pada grup ke-k
- $\bar{\pi}_k$: $\sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$ yaitu rata-rata taksiran peluang
- g : Jumlah grup (kombinasi kategori dalam model serentak)
- n'_k : Banyak observasi pada grup ke-k

Nilai Kritis :

$X^2_{(\alpha, g-2)}$: nilai kritis diperoleh dari distribusi Chi Kuadrat dengan derajat bebas $g-2$.

Kriteria Uji :

Kriteria pengujian mengambil taraf nyata α maka tolak H_0 jika $\hat{C} > X^2_{(\alpha, g-2)}$ atau $pvalue < \alpha$.

4. Ketepatan Klasifikasi

Keutamaan model dapat diukur menggunakan ketepatan klasifikasi. Ukuran untuk melihat peluang gagal dalam ketepatan klasifikasi adalah *Apparent Error Rate* (APER). APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson & Winchern, 2007). Mengukur kesalahan klasifikasi dapat menggunakan *confusion matrix*. Tabel *confusion matrix* untuk menghitung tingkat kesalahan klasifikasi atau APER.

Tabel 1. *Confusion Matrix* Observasi Aktual dan hasil Prediksi Model

Kelompok Aktual	Kelompok Prediksi		Total
	π_1	π_2	
π_1	n_{11}	$n_{1M} = n_1 n_{1c}$	n_1
π_2	$n_{2M} = n_2 n_{2c}$	n_{2c}	n_2

Keterangan:

- n_{1c} = jumlah prediksi π_1 yang tepat diklasifikasikan ke π_1
- n_{1M} = jumlah prediksi π_1 yang tepat diklasifikasikan ke π_2
- n_{2c} = jumlah prediksi π_2 yang tepat diklasifikasikan ke π_2
- n_{2M} = jumlah prediksi π_2 yang tepat diklasifikasikan ke π_1
- n_1 = jumlah item yang masuk dalam kelompok 1
- n_2 = jumlah item yang masuk dalam kelompok 2

Nilai APER dapat diperoleh berdasarkan persamaan sebagai berikut:

$$APER = \left(\frac{n_{1M} + n_{2M}}{n_1 + n_2} \right) \times 100\%$$

$$Ketepatan\ Klasifikasi = 1 - APER$$

Penyakit Jantung Koroner

Penyakit Jantung Koroner (PJK) merupakan kondisi dimana pembuluh darah jantung (arteri koroner) tersumbat oleh timbunan lemak yang berlapis-lapis. Penderita penyakit jantung koroner banyak didapatkan karena adanya faktor-faktor risiko. Faktor risiko penyakit jantung koroner dapat dibagi menjadi dua golongan besar, yaitu faktor risiko yang dapat dikurangi, diperbaiki dan diubah yaitu merokok, kolesterol, diabetes, glukosa dan IMT, dan faktor risiko yang bersifat alami atau tidak dapat diubah yaitu usia dan jenis kelamin.

3. Hasil dan Pembahasan

Regresi Logit Biner

Analisis regresi logit biner dilakukakn dengan meregresikan variabel risiko penyakit jantung

koroner (Y) dengan variabel *Gender* (X_1), *Usia* (X_2), *Perokok Aktif* (X_3), *Diabetes* (X_4), *Kolesterol* (X_5), *IMT* (X_6) dan *Glukosan* (X_7). Kemudian akan didapatkan model dari regresi logit biner yang akan diuji dengan uji serentak dan parsial.

Pengujian Parameter

1. Uji Serentak

Tabel 2. Uji Serentak Model Regresi Logit Biner

G	Chi-Square	Keputusan
310,357	14,06714	Tolak H_0

Pada tabel di atas menunjukkan bahwa uji serentak menghasilkan H_0 di tolak karena $G > \chi^2_{(0,05,7)}$ yaitu $310,357 > 14,06714$, sehingga dapat disimpulkan bahwa minimal terdapat satu variabel prediktor yang signifikan mempengaruhi risiko penyakit jantung koroner.

2. Uji Parsial

Tabel 3. Uji Parsial Model Regresi Logit Biner

Variabel	Estimate	Std.Error	Wald	Pvalue	Keputusan
X1	0,549853	0,098571	31,11679	2,43e-08	Tolak H_0
X2	0,077790	0,005965	170,06958	< 2e-16	Tolak H_0
X3	0,399289	0,101434	15,49557	8,27e-05	Tolak H_0
X4	0,040137	0,305923	0,017213	0,895617	Terima H_0
X5	0,003445	0,001050	10,76465	0,001033	Tolak H_0
X6	0,039169	0,011451	11,70033	0,000625	Tolak H_0
X7	0,007915	0,002174	13,25509	0,000272	Tolak H_0

Berdasarkan tabel di atas dapat diketahui bahwa variabel gender, usia, perokok aktif, kolesterol, IMT dan glukosa signifikan berpengaruh terhadap variabel risiko penyakit jantung koroner. Sedangkan variabel diabetes tidak signifikan berpengaruh terhadap variabel risiko penyakit jantung koroner, hal ini dapat ditunjukkan dengan nilai $W_j < \chi^2_{(0,05;1)} = 3.84$ atau nilai $pvalue > \alpha = 0,05$.

Uji Kesesuaian Model

Berdasarkan uji Hosmer dan Lemeshow diperoleh nilai $pvalue > \alpha$ yaitu $0,05236 > 0,05$ sehingga H_0 diterima. Maka, model regresi logistik biner yang terbentuk sesuai atau tidak ada perbedaan antara observasi dengan kemungkinan hasil prediksi.

Model terbaik Regresi Logit Biner

Berdasarkan pengujian di atas, didapatkan hasil bahwa variabel diabetes tidak berpengaruh signifikan terhadap variabel risiko penyakit jantung koroner. Karena ada satu variabel yang tidak signifikan, maka dilakukan pengujian ulang, sehingga diperoleh model terbaik untuk regresi logit biner sebagai berikut:

$$\pi(x) = \frac{\exp(-8,734 + 0,55x_1 + 0,078x_2 + 0,399x_3 + 0,003x_5 + 0,039x_6 + 0,008x_7)}{1 + \exp(-8,734 + 0,55x_1 + 0,078x_2 + 0,399x_3 + 0,003x_5 + 0,039x_6 + 0,008x_7)}$$

Ketepatan Klasifikasi Regresi Logit Biner

Tabel 4. Confusion Matrix Model Regresi Logit Biner

Observasi	Prediksi		Total
	Negative PJK	Positive PJK	
Negative PJK	3.228	7	3.235
Positive PJK	567	18	585
Total	3.795	25	3.820

Berdasarkan Tabel 4.6 dapat dilihat bahwa terdapat 3.228 orang yang negatif berisiko penyakit jantung koroner diprediksi benar oleh model dan terdapat 18 orang yang positif berisiko penyakit jantung koroner diprediksi benar oleh model. Tingkat kesalahan klasifikasi pada persamaan (2.23) dan ketepatan klasifikasi pada persamaan (2.24) dapat dilakukan perhitungan sebagai berikut.

$$APER = \left(\frac{7 + 567}{3.820} \right) \times 100\% = 15,03\%$$

$$\text{Ketepatan klasifikasi} = 1 - 15,03\% = 84,92\%$$

Maka dari model regresi logistik biner yang terbentuk menghasilkan nilai akurasi atau ketepatan klasifikasi yang tinggi yaitu sebesar 84,92% sehingga regresi logistik biner untuk data risiko penyakit jantung koroner baik untuk digunakan.

Regresi Probit Biner

Pengujian Parameter

1. Uji Serentak

Tabel 5. Uji Serentak Regresi Probit Biner

G	Chi-Square	Keputusan
311,7490513	14,06714	Tolak H_0

Pada tabel di atas menunjukkan bahwa uji serentak menghasilkan H_0 di tolak karena $G > \chi^2_{(0,05,7)}$ yaitu $311,749 > 14,06714$, sehingga dapat disimpulkan bahwa minimal terdapat satu variabel prediktor yang signifikan mempengaruhi risiko penyakit jantung koroner.

2. Uji Parsial

Tabel 6. Uji Parsial Regresi Probit Biner

Variabel	Estimate	Std.Error	Wald	Pvalue	Keputusan
X1	0,2963561	0,053951	30,173712	3,95e-08	Tolak H_0
X2	0,043886	0,003261	181,11332	< 2e-16	Tolak H_0
X3	0,227765	0,003261	16,66627	4,47e-05	Tolak H_0
X4	0,047728	0,055791	0,731843	0,788366	Terima H_0
X5	0,001766	0,000588	9,020419	0,002656	Tolak H_0
X6	0,020468	0,006407	10,20567	0,001400	Tolak H_0
X7	0,0044003	0,001247	12,45179	0,000418	Tolak H_0

Berdasarkan tabel di atas dapat diketahui bahwa variabel gender, usia, perokok aktif, kolesterol, IMT dan glukosa signifikan berpengaruh terhadap variabel risiko penyakit jantung koroner. Sedangkan variabel diabetes tidak signifikan berpengaruh terhadap variabel risiko penyakit jantung koroner, hal ini dapat ditunjukkan dengan nilai $W_j < \chi^2_{(0,05;1)} = 3,84$ atau nilai $pvalue > \alpha = 0,05$.

Uji Kesesuaian Model

Berdasarkan uji Hosmer dan Lemeshow diperoleh nilai $pvalue > \alpha$ yaitu $0,07031 > 0,05$ sehingga H_0 diterima. Maka, model regresi probit biner yang terbentuk sesuai atau tidak ada perbedaan antara observasi dengan kemungkinan hasil prediksi.

Model Terbaik Regresi Probit Biner

Berdasarkan pengujian di atas, didapatkan hasil bahwa variabel diabetes tidak berpengaruh signifikan terhadap variabel risiko penyakit jantung koroner. Karena ada satu variabel yang tidak signifikan, maka dilakukan pengujian ulang, sehingga diperoleh model terbaik untuk regresi probit biner sebagai berikut:

$$P(Y = 0) = \Phi(-4,844 + 0,296x_1 + 0,043x_2 + 0,228x_3 + 0,002x_5 + 0,02x_6 + 0,005x_7)$$

$$P(Y = 1) = 1 - \Phi(-4,844 + 0,296x_1 + 0,043x_2 + 0,228x_3 + 0,002x_5 + 0,02x_6 + 0,005x_7)$$

Ketepatan Klasifikasi Regresi Probit Biner

Tabel 7. *Confusion Matrix* Model Probit Biner

Observasi	Prediksi		Total
	Negative PJK	Positive PJK	
Negative PJK	3.229	6	3.235
Positive PJK	570	15	585
Total	3.799	21	3.820

Berdasarkan Tabel 4.7 dapat dilihat bahwa terdapat 3.229 orang yang negatif berisiko penyakit jantung koroner diprediksi benar oleh model dan terdapat 15 orang yang positif berisiko penyakit jantung koroner diprediksi benar oleh model. Tingkat kesalahan klasifikasi pada persamaan (2.22) dan ketepatan klasifikasi pada persamaan (2.23) dapat dilakukan perhitungan sebagai berikut.

$$APER = \left(\frac{6 + 570}{3820} \right) \times 100\% = 15,08\%$$

$$\text{Ketepatan klasifikasi} = 1 - 15,08\% = 84,92\%$$

Maka dari model regresi logistik biner yang terbentuk menghasilkan nilai akurasi atau ketepatan klasifikasi yang tinggi yaitu sebesar 84,92% sehingga regresi probit biner untuk data risiko penyakit jantung koroner baik untuk digunakan.

Perbandingan Regresi Model Logit Biner dan Model Probit Biner

Tabel 8. Perbandingan Ketepatan Klasifikasi

Model	Ketepatan Klasifikasi (%)	
	APER	AKURASI
Regresi Logistik Biner	15,05	84,95
Regresi Probit Biner	15,08	84,92

Berdasarkan tabel di atas, nilai akurasi regresi logistik biner sebesar 84,95% dengan nilai APER 15,05% sedangkan nilai akurasi regresi probit biner sebesar 84,92% dengan nilai APER 15,08%. Hal ini menunjukkan bahwa menurut Johnson dan Winchern, pemodelan status risiko penyakit jantung koroner menggunakan regresi logistik biner dan regresi probit biner menghasilkan nilai akurasi klasifikasi yang relatif sama karena perbedaan nilai akurasi kedua metode tersebut sangat kecil atau untuk kedua model sama baiknya digunakan untuk mengklasifikasikan status risiko penyakit jantung koroner.

4. Kesimpulan

Berdasarkan hasil dan pembahasan mengenai risiko penyakit jantung koroner di Kota Framingham Amerika, dengan menggunakan regresi logit biner dan probit biner, diberikan kesimpulan sebagai berikut:

Risiko penyakit jantung koroner di Kota Framingham Amerika, dipengaruhi oleh faktor jenis kelamin, usia, perokok aktif, kolesterol, IMT dan glukosa.

Dari hasil perbandingan ketepatan klasifikasi antara metode regresi logit biner dan metode regresi probit biner, kedua metode relatif sama atau kedua metode sama baiknya digunakan untuk mengklasifikasikan risiko penyakit jantung koroner di Kota Framingham Amerika.

5. Saran

Adapun saran untuk penelitian selanjutnya adalah selain menggunakan ketepatan klasifikasi, banyak cara untuk menentukan perbandingan metode regresi model logit biner dan regresi model probit biner, diantaranya yaitu membandingkan AIC atau membandingkan Uji kelayakan model (*Goodness of Fit*) kedua model.

Daftar Pustaka

- [1] Agresti, A. (2002). *Categorical Data Analysis Second Edition*. John Willey and Sons. New York.
- [2] Greene, W. H. (2008). *Econometrics Analysis Sixth Edition*. Prentice Hall: New Jersey.
- [3] Gujarati, D. N. (2004). *Basic Econometrics*. New York: McGraw-Hill Companies.
- [4] Hosmer, D. W., dan Lemeshow. (2000). *Applied Logistic Regression*. USA: John Willey and Sons.
- [5] Johnson, R. A., dan Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice Hall: New Jersey.
- [6] Kleinbaum, D., dan Klein, M. (2002). *Logistic Regression*. New York: Spinger Verlag.
- [7] Setiawan, F. H., Rahmawati R., dan Suparti. (2014). *Ketepatan Klasifikasi Keikutsertaan Keluarga Berencana Menggunakan Regresi Logistik Biner dan Regresi Probit Biner*. Semarang: Jurnal Gaussian.