

KETEPATAN KLASIFIKASI KEIKUTSERTAAN KELUARGA BERENCANA MENGUNAKAN REGRESI LOGISTIK BINER DAN REGRESI PROBIT BINER (Studi Kasus di Kabupaten Semarang Tahun 2014)

Fajar Heru Setiawan¹, Rita Rahmawati², Suparti³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

fajarherusetiawan23@gmail.com,

ABSTRACT

Population growth in Indonesia has increased each year. According to the population census conducted by National Statistics Bureau in 2010, Indonesia's population reached 237,6 million. Therefore, to control the population growth rate, government hold Keluarga Berencana (KB) or family planning program for couples in the childbearing age. The aim of this thesis which analyze the classification of couples in the childbearing age who follow family planning program, is to reduce the birth rate. So that, population can be controlled. The data used in this study is a Semarang Regency updated family data in 2014 that conducted Nasional Population and Family Planning Bureau. From the data, a binary logistic regression model and binary probit regression will be obtained, also classification accuracy will be obtained from each of these models. The analysis showed that the Binary Logistic Regression method produces a classification accuracy of 69,0% with 31,0% classification error. While, Probit Binary Regression method produces a classification accuracy of 68,4% with 31,6% misclassification. Binary Logistic Regression and Binary Logistic Regression method have a differences classification accuracy was very small then both are relative similar for analyze the classification family planning in Semarang Regency.

Keywords: Keluarga Berencana (KB), Binary Logistic Regression, Binary Probit Regression, Classification, Confusion

1. PENDAHULUAN

1.1. Latar Belakang

Menurut sensus penduduk oleh Badan Pusat Statistik (BPS) tahun 2010 jumlah penduduk Indonesia adalah 237.556.363 orang, yang terdiri dari 119.507.580 laki-laki dan 118.048.783 perempuan. Diketahui juga dari hasil sensus penduduk 2010 bahwa laju pertumbuhan penduduk sebesar 1,49 persen per tahun. Penduduk Indonesia terus bertambah dari waktu ke waktu, dari awal dilakukannya sensus pada tahun 1930 sampai sensus tahun 2010 yang mencapai 237,6 juta jiwa (BPS, 2010).

Fertilitas merupakan salah satu faktor yang mempengaruhi pertumbuhan penduduk selain mortalitas dan migrasi karena fertilitas merupakan penyumbang tingginya angka kelahiran. Tingkat fertilitas yang tinggi merupakan masalah yang serius bagi suatu negara tak terkecuali Indonesia. Pertumbuhan penduduk yang tinggi mengakibatkan timbulnya berbagai masalah mulai dari masalah sosial, ekonomi, kesehatan, pendidikan sampai dengan permasalahan kerusakan lingkungan. Pemerintah telah melakukan upaya dan berbagai solusi untuk mengatasi masalah yang timbul akibat laju pertumbuhan penduduk salah satunya adalah mengadakan program Keluarga Berencana (KB) yang ditujukan kepada Pasangan Usia Subur (PUS) dengan cara pemakaian alat kontrasepsi (BKKBN, 2014)

Keluarga Berencana adalah suatu usaha yang mengatur banyaknya kelahiran sedemikian rupa sehingga bagi ibu maupun bayinya dan bagi ayah serta keluarganya atau masyarakat yang bersangkutan tidak akan menimbulkan kerugian sebagai akibat langsung dari kelahiran tersebut. Tujuan dari KB adalah meningkatkan kesejahteraan ibu dan anak dalam rangka mewujudkan NKKBS (Norma Keluarga Kecil Bahagia Sejahtera) yang

menjadi dasar terwujudnya masyarakat yang sejahtera dengan mengendalikan kelahiran sekaligus menjamin terkendalinya pertumbuhan penduduk (Irianto, 2014).

Beberapa penelitian tentang KB yang telah dilakukan sebelumnya, menyebutkan bahwa variabel yang berpengaruh terhadap keikutsertaan KB pasangan usia subur adalah umur ibu, umur anak terakhir dan pernah tidaknya mendapatkan penyuluhan terhadap keluarga berencana dari pihak berwenang. Untuk menganalisis pasangan usia subur mengikuti KB atau tidak dapat menggunakan metode regresi logistik biner atau regresi probit biner karena untuk variabel dependennya (keikutsertaan KB) hanya memiliki dua nilai dan bersifat kualitatif. Menurut Agresti (2002) salah satu cara yang digunakan untuk menganalisis hubungan antara suatu variabel dependen dan beberapa variabel independen, dimana variabel dependennya berupa data kualitatif dikotomi yang bernilai 1 atau 0 adalah analisis regresi logistik biner. Selain Regresi Logistik Biner untuk menganalisis variabel dependen yang dikotomi dapat juga digunakan regresi probit biner. Regresi probit biner menggunakan bilangan biner (peubah *dummy*) sebagai variabel dependennya yang berbentuk kualitatif. Kedua metode dapat menganalisis pula nilai ketepatan klasifikasi. Berdasarkan uraian tersebut penulis memilih judul “Ketepatan Klasifikasi Keikutsertaan Keluarga Berencana Menggunakan Regresi Logistik Biner dan Regresi Probit Biner (Studi Kasus di Kabupaten Semarang Tahun 2014)”.

1.2. Tujuan Penelitian

Berdasarkan latar belakang tersebut, tujuan penelitian ini adalah:

1. Menentukan faktor-faktor yang mempengaruhi keikutsertaan program Keluarga Berencana.
2. Menentukan ketepatan klasifikasi peserta keluarga berencana berdasarkan analisis regresi logistik biner.
3. Menentukan ketepatan klasifikasi peserta keluarga berencana berdasarkan analisis regresi probit biner.
4. Membandingkan ketepatan klasifikasi berdasarkan analisis regresi logistik biner dan regresi probit biner.

2. TINJAUAN PUSTAKA

2.1. Regresi Logistik Biner

Menurut Hosmer dan Lemeshow (2000) regresi logistik merupakan metode yang menghubungkan antara variabel dependen yang bersifat kategorik dengan variabel independen yang bersifat kategorik atau kontinu. Model regresi logistik dengan variabel dependen yang terdiri dari dua kategori disebut model regresi logistik biner (dikotomus).

Model regresi logistik biner bertujuan mendapatkan pola hubungan antara variabel X dengan $\pi(x_i)$ dimana $\pi(x_i)$ adalah probabilitas kejadian yang diakibatkan variabel X. Sehingga bentuk model regresi logistiknya adalah sebagai berikut:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}$$

$$\text{Sedangkan logit dari } \pi(x_i) \text{ adalah } g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

Jika $g(x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ maka $g(x_i)$ merupakan fungsi hubungan dari model regresi logistik yang disebut fungsi hubungan logit.

Pada keikutsertaan Keluarga Berencana (KB) dalam penelitian ini, untuk menentukan prediksi peluang pasangan usia subur mengikuti KB atau pasangan usia subur

tidak mengikuti KB menggunakan *cut value* 0,5 atau $\pi(x_i) = 0,5$. Artinya, jika nilai peluang $\pi(x_i)$ yang dihasilkan $< 0,5$ maka pasangan usia subur dikategorikan tidak mengikuti KB atau masuk dalam kategori kelas 0 dan jika nilai peluang $\pi(x_i)$ yang dihasilkan $\geq 0,5$ maka pasangan usia subur dikategorikan mengikuti KB atau masuk dalam kategori kelas 1. Berdasarkan pernyataan tersebut dapat dibentuk persamaan berikut

$$y_i = \begin{cases} 1 & , \text{jika } \pi(x_i) \geq 0,5 \\ 0 & , \text{jika } \pi(x_i) < 0,5 \end{cases}$$

2.2. Estimasi Parameter

Menurut Agresti (2002) untuk menentukan estimasi parameter regresi logistik biner dapat digunakan metode *Maximum Likelihood* (kemungkinan maksimum) yang membutuhkan turunan pertama dan turunan kedua dari fungsi Likelihood. Variabel dependen y_i berdistribusi bernoulli karena variabel y_i hanya memiliki dua nilai yang mungkin yaitu “sukses” dan “gagal” dimana dapat dinotasikan dengan angka 1 dan 0. Jika 1 merupakan kejadian “sukses” maka fungsi kepadatan peluangnya adalah:

$$p(y_i = 1) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Menurut Hosmer dan Lemeshow (2000) untuk setiap pengamatan sebanyak n diasumsikan independen maka fungsi likelihoodnya merupakan perkalian antara masing-masing fungsi densitasnya yaitu:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Pada prinsipnya, fungsi maksimum likelihood menggunakan estimasi nilai β untuk memaksimalkan fungsi likelihood. Dan hasil fungsi logaritma likelihoodnya adalah:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))]$$

Hosmer dan Lemeshow (2000) menyatakan untuk memperoleh nilai β , yaitu dengan cara memaksimalkan $L(\beta)$ dan mendiferensialkan $L(\beta)$ terhadap β_0 dan β_j dengan $j = 1, 2, \dots, p$ dan menyamadengkan dengan 0, persamaan ini dapat ditulis dalam bentuk sebagai berikut:

$$\sum_{i=1}^n (y_i - \pi(x_i)) = 0 \text{ dan } \sum_{i=1}^n (x_i [y_i - \pi(x_i)]) = 0$$

Di dalam Hosmer dan Lemeshow (2000) metode untuk mencari estimasi varian kovarian dari koefisien estimasi mengikuti teori yang dikembangkan dengan teori estimasi likelihood. Teori tersebut menyatakan bahwa estimator diperoleh dari matriks turunan parsial kedua dari fungsi log likelihood. Menurut Hosmer dan Lemeshow (2000) turunan parsial memiliki bentuk umum berikut:

$$\frac{\partial^2 L(\beta)}{(\partial \beta_j)^2} = - \sum_{i=1}^n (x_{ji})^2 \pi(x_i) [1 - \pi(x_i)] \text{ dan } \frac{\partial^2 L(\beta)}{\partial \beta_u \partial \beta_j} = - \sum_{i=1}^n x_{ui} x_{ji} \pi(x_i) [1 - \pi(x_i)]$$

2.3. Regresi Probit Biner

Model probit dikenalkan pertama kali oleh Chester Ittner Bliss pada tahun 1935. Model probit disebut juga model normit yang menggunakan bilangan biner (variabel *dummy*) sebagai variabel dependennya yang berbentuk kualitatif, sedangkan variabel independennya berbentuk kualitatif atau kuantitatif. Variabel *dummy* yang dimaksud di sini adalah jenis variabel diskrit yang mempunyai dua nilai. Model ini adalah model yang non

linier karena variabel bebasnya linier terhadap probabilitas y_i bukan linier terhadap y_i dan mengasumsikan galat ε_i berdistribusi normal dengan mean 0 dan varian σ^2 . Sehingga untuk menjelaskan model probit ini digunakan fungsi distribusi kumulatif (CDF) dari distribusi normal.

Misalkan ada variabel respon y_i^* yang menunjukkan suatu kejadian yang sukses dan X_1, X_2, \dots, X_p adalah variabel bebas. Variabel respon ini dipengaruhi oleh beberapa karakteristik individu dan kondisi lingkungan sehingga persamaan y_i^* dapat dituliskan sebagai (Greene, 2000):

$$y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$$

2.4. Estimasi Parameter Regresi Probit

Metode *Maximum Likelihood Estimator* (MLE) digunakan untuk menduga parameter-parameter dari suatu model yang diketahui distribusinya. Pada model probit, y_i^* tidak dapat diamati tetapi tindakan individu dapat diamati bila y_i^* melewati batas tertentu. Jika $y_i^* > 0$ maka $y_i = 1$ dan jika $y_i^* \leq 0$ maka $y_i = 0$. Dari $y_i = 0$ diperoleh $P(y_i = 0) = 1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta})$ dan untuk $y_i = 1$ diperoleh $P(y_i = 1) = \Phi(\mathbf{x}_i^T \boldsymbol{\beta})$ dengan $\Phi(\cdot)$ adalah fungsi distribusi normal kumulatif dan $\Phi(\mathbf{x}_i^T \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$ ketika berdistribusi normal dengan rata-rata 0 dan varian 1.

Model dengan peluang sukses $\Phi(\mathbf{x}_i^T \boldsymbol{\beta})$ dengan n pengamatan yang saling bebas didapatkan fungsi likelihoodnya sebagai berikut:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \Phi(\mathbf{x}_i^T \boldsymbol{\beta})^{y_i} [1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta})]^{1-y_i}$$

Dengan melogaritmakan fungsi likelihoodnya diperoleh:

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) + \sum_{i=1}^n (1 - y_i) \ln (1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta}))$$

Menurut Greene (2000) untuk memaksimumkan \ln fungsi likelihoodnya menurunkan $L(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ dan disamadengankan nol dihasilkan:

$$\sum_{i=1}^n \lambda_i \mathbf{x}_i = 0 \text{ dengan } \lambda_i = \frac{q_i \phi(q_i \mathbf{x}_i^T \boldsymbol{\beta})}{\Phi(q_i \mathbf{x}_i^T \boldsymbol{\beta})}, \quad q_i = 2y_i - 1$$

Persamaan $\sum_{i=1}^n \lambda_i \mathbf{x}_i$ bukan merupakan fungsi linier dalam $\boldsymbol{\beta}$ dan hasilnya secara analitik tidak dapat langsung diketahui sehingga taksiran $\boldsymbol{\beta}$ dicari dengan menggunakan metode numerik. Metode yang digunakan adalah metode iterasi Newton-Raphson. Metode ini membutuhkan turunan parsial kedua dari fungsi likelihood. Turunan keduanya adalah:

$$-\sum_{i=1}^n (\lambda_i (\boldsymbol{\beta}^T \mathbf{x}_i + \lambda_i)) \mathbf{x}_i \mathbf{x}_i^T$$

Dalam menganalisis metode regresi logistik biner dan metode regresi probit biner ada beberapa uji yang dilakukan yaitu uji serentak untuk mengetahui apakah semua variabel independen berpengaruh terhadap variabel dependen atau tidak, uji individu untuk

mengetahui variabel independen mana saja yang berpengaruh terhadap variabel dependen, uji kesesuaian model untuk mengetahui apakah model yang terbentuk sesuai atau tidak.

2.5. Uji Serentak

Hipotesis yang digunakan adalah:

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

$$H_1 : \text{paling sedikit ada satu } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

$$\text{Statistik uji } G \text{ atau Likelihood Ratio Test } G = -2 \ln \left[\frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}} \right]$$

Daerah penolakan : Tolak H_0 jika $G > \chi^2_{(p,\alpha)}$ atau H_0 ditolak bila nilai $p\text{-value} < \alpha$.

2.6. Uji Partial

Hipotesis yang digunakan adalah:

$$H_0 : \beta_j = 0 \text{ dengan } j = 1, 2, \dots, p$$

$$H_1 : \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

$$\text{Statistik uji Wald: } W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$

Daerah penolakan: Tolak H_0 jika $W_j > \chi^2_{(\alpha,1)}$ atau nilai $p\text{-value} < \alpha$.

2.7. Uji Kesesuaian Model

2.7.1. Uji Hosmer dan Lemeshow

Hipotesis yang digunakan adalah:

H_0 : Model sesuai

H_1 : Model tidak sesuai

$$\text{Statistik uji Hosmer dan Lemeshow test: } \hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Daerah penolakan : Tolak H_0 jika $\hat{C} > \chi^2_{(\alpha, g-2)}$ atau H_0 ditolak bila nilai $p\text{-value} < \alpha$.

2.7.2. Uji Pearson Chi-Square

Hipotesis yang digunakan adalah:

H_0 : Model sesuai

H_1 : Model tidak sesuai

$$\text{Statistik uji dari Pearson Chi-Square test: } \chi^2 = \sum_j r_j^2$$

Daerah penolakan : Tolak H_0 jika $\chi^2 > \chi^2_{(\alpha, J-p-1)}$ atau H_0 ditolak bila nilai $p\text{-value} < \alpha$.

2.8. Ketepatan Klasifikasi

Menurut Johnson dan Wichern (1992) prosedur klasifikasi yang dipakai pada regresi logistik adalah *apparent error rate* (APER). Nilai APER menyatakan proporsi sampel

yang salah diklasifikasikan oleh fungsi klasifikasi. Kesalahan klasifikasi dapat dilihat pada tabel matriks konfusi berikut:

Tabel 1. Matriks Konfusi

Hasil	Taksiran	
Observasi	y_1	y_2
y_1	n_{11}	n_{12}
y_2	n_{21}	n_{22}

Maka nilai APER dapat dihitung dengan rumus: $APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$

3. METODE PENELITIAN

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yaitu data Pemutakhiran Data Keluarga (MDK) Kabupaten Semarang tahun 2014. Variabel yang digunakan pada penelitian ini terdiri atas variabel dependen (Y) yaitu keikutsertaan KB yang dibagi menjadi ikut KB (Y=1) dan tidak ikut KB (Y=0). Variabel independen umur ibu (X1) dibagi menjadi (≤ 35 tahun dan > 35 tahun), jumlah anak (X2) dibagi menjadi (≤ 2 anak dan > 2 anak), pendidikan ayah (X3) dan pendidikan ibu (X4) masing-masing dibagi menjadi (pendidikan rendah, pendidikan sedang dan pendidikan tinggi), dan tingkat kesejahteraan (X5) yang dibagi menjadi (pra sejahtera, keluarga sejahtera I, keluarga sejahtera II, keluarga sejahtera III, keluarga sejahtera III+).

Tahapan analisis yang dilakukan adalah pembentukan model regresi logistik biner dan regresi probit biner, setelah didapatkan model awal masing-masing metode kemudian melakukan uji serentak menggunakan uji Rasio Likelihood dan uji individu menggunakan uji Wald. Jika kedua uji terpenuhi kemudian melakukan uji kesesuaian model, untuk metode regresi logistik biner menggunakan uji Hosmer dan Lemeshow sedangkan untuk metode regresi probit biner menggunakan uji Pearson Chi-Square. Setelah semua uji terpenuhi, maka didapatkan model akhir dari masing-masing metode. Model akhir kedua metode kemudian dihitung nilai ketepatan klasifikasinya dan dibandingkan. Hasil dari perbandingan nilai ketepatan klasifikasi akan didapatkan kesimpulan metode mana yang memiliki model terbaik.

4. HASIL DAN PEMBAHASAN

4.1. Regresi Logistik Biner

Analisis regresi logistik biner dilakukan dengan meregresikan variabel keikutsertaan KB (Y) dengan variabel umur ibu (X₁), variabel jumlah anak (X₂), variabel pendidikan ayah (X₃), variabel pendidikan ibu (X₄) dan variabel tingkat kesejahteraan (X₅). Kemudian akan didapatkan model dari regresi logistik biner yang akan diuji dengan uji serentak dan individu.

4.1.1. Uji Rasio Likelihood

Tabel 2. Uji Rasio Likelihood

Uji Ke-	G	Nilai Tabel Chi-Square	Keputusan
1	74,456	$\chi^2_{(10;0,05)} 18,307$	H ₀ ditolak
2	65,519	$\chi^2_{(4;0,05)} 9,488$	H ₀ ditolak

Berdasarkan Tabel 2 diketahui bahwa uji rasio likelihood menghasilkan keputusan H_0 di tolak karena nilai $G > \chi^2_{(p;0,05)}$, sehingga dapat disimpulkan bahwa variabel independen berpengaruh nyata secara serentak.

4.1.2. Uji Wald

Tabel 3. Uji Wald Pertama

Variabel	Wald	<i>P-value</i>	Keputusan
X1	16,049	0,000	H_0 ditolak
X2	42,353	0,000	H_0 ditolak
X3(1)	1,950	0,163	H_0 diterima
X3(2)	3,102	0,078	H_0 diterima
X4(1)	12,204	0,000	H_0 ditolak
X4(2)	2,225	0,136	H_0 diterima
X5(1)	1,591	0,207	H_0 diterima
X5(2)	0,172	0,678	H_0 diterima
X5(3)	2,292	0,130	H_0 diterima
X5(4)	2,213	0,137	H_0 diterima
Constant	1,247	0,264	

Tabel 4. Uji Wald Kedua

Variabel	Wald	<i>P-value</i>	Keputusan
X1	16,057	0,000	H_0 ditolak
X2	43,242	0,000	H_0 ditolak
X4(1)	9,644	0,002	H_0 ditolak
X4(2)	1,113	0,291	H_0 diterima
Constant	4,793	0,029	

Keterangan:

X1 : Umur Ibu

X2 : Jumlah Anak

X3 : Pendidikan Ayah

X4 : Pendidikan Ibu

X5 : Tingkat Kesejahteraan

Berdasarkan Tabel 3, didapatkan hasil bahwa variabel pendidikan ayah dan tingkat kesejahteraan tidak signifikan terhadap variabel keikutsertaan KB, hal ini ditunjukkan dengan nilai $W_j < \chi^2_{(1;0,05)} = 3,84$ atau nilai $p\text{-value} > \alpha = 0,05$. Karena ada dua variabel yang tidak signifikan maka akan dibentuk model baru yang diuji lagi dengan uji rasio likelihood ke-2 yang nilainya pada baris kedua dalam Tabel 2 dan uji wald kedua pada Tabel 4. Dari Tabel 4 didapatkan kesimpulan semua variabel yaitu umur ibu, jumlah anak dan pendidikan ibu signifikan yang ditunjukkan dengan $W_j > \chi^2_{(1;0,05)} = 3,84$ atau nilai $p\text{-value} > \alpha = 0,05$.

4.1.3. Uji Kesesuaian Model

Berdasarkan uji kesesuaian model diperoleh nilai $\hat{C} < \chi^2_{(0,05;7)}$ yaitu $3,211 < 14,067$ atau $p\text{-value} > \alpha$ yaitu $0,865 > 0,05$ sehingga H_0 diterima. Jadi, model regresi logistik biner yang terbentuk sesuai atau tidak ada perbedaan antara observasi dengan kemungkinan hasil prediksi.

4.1.4. Model Akhir

Setelah dilakukan uji signifikansi terhadap model, baik secara keseluruhan maupun individual serta dilakukan uji kesesuaian model maka diperoleh model akhir sebagai berikut:

$$\pi(x_i) = \frac{\exp(-0,668 - 1,108 + 1,827 + 1,025(1) + 0,363(2))}{1 + \exp(-0,668 - 1,108 + 1,827 + 1,025(1) + 0,363(2))}$$

4.2. Ketepatan Klasifikasi

Dari hasil perhitungan nilai $\pi(x_i)$, jika untuk nilai $\pi(x_i) < 0,5$ pasangan usia subur dikategorikan tidak mengikuti KB sedangkan untuk nilai $\pi(x_i) \geq 0,5$ pasangan usia subur dikategorikan mengikuti KB. Sehingga dapat dibentuk tabel berikut:

Tabel 5. Matriks Konfusi Klasifikasi Metode Regresi Logistik Biner

Observasi	Prediksi		Akurasi (%)
	Tidak ikut KB	Ikut KB	
Tidak ikut KB	95	77	55,2
Ikut KB	45	176	79,6
Total			69,0

Berdasarkan Tabel 5, dapat dihitung nilai ketepatan klasifikasinya adalah 69% yang didapat dari membandingkan jumlah dari pasangan usia subur yang tepat diklasifikasikan dengan jumlah seluruh observasi dan nilai kesalahan klasifikasinya (APER) adalah 31%.

4.3. Analisis Regresi Probit

Analisis regresi probit biner dilakukan dengan meregresikan variabel keikutsertaan KB (Y) dengan variabel umur ibu (X_1), variabel jumlah anak (X_2), variabel pendidikan ayah (X_3), variabel pendidikan ibu (X_4) dan variabel tingkat kesejahteraan (X_5). Kemudian akan didapatkan model dari regresi probit biner yang akan diuji dengan uji serentak dan individu.

4.3.1. Uji Rasio Likelihood

Tabel 6. Uji Rasio Likelihood

Uji Ke-	G	Nilai Tabel Chi-Square	Keputusan
1	79,955	$\chi^2_{(10;0,05)} 18,307$	H_0 ditolak
2	68,420	$\chi^2_{(4;0,05)} 9,488$	H_0 ditolak

Berdasarkan Tabel 6 diketahui bahwa uji rasio likelihood menghasilkan keputusan H_0 di tolak karena nilai $G > \chi^2_{(p;0,05)}$, sehingga dapat disimpulkan bahwa variabel independen berpengaruh nyata secara serentak.

4.3.2. Uji Wald

Tabel 7. Uji Wald Pertama

Variabel	Wald	<i>p-value</i>	Keputusan
X1	16,679	0,000	H_0 ditolak
X2	44,192	0,000	H_0 ditolak
X3(1)	1,567	0,211	H_0 diterima
X3(2)	2,609	0,106	H_0 diterima
X4(1)	11,943	0,001	H_0 ditolak
X4(2)	1,797	0,180	H_0 diterima
X5(1)	1,552	0,213	H_0 diterima
X5(2)	0,230	0,631	H_0 diterima
X5(3)	2,116	0,146	H_0 diterima
X5(4)	2,064	0,151	H_0 diterima
Constant	0,016	0,900	

Tabel 8. Uji Wald Kedua

Variabel	Wald	<i>p-value</i>	Keputusan
X1	16,269	0,000	H_0 ditolak
X2	45,949	0,000	H_0 ditolak
X4(1)	10,027	0,002	H_0 ditolak
X4(2)	1,084	0,298	H_0 diterima
Constant	0,005	0,946	

Keterangan:

X1 : Umur Ibu

X2 : Jumlah Anak

X3 : Pendidikan Ayah
 X4 : Pendidikan Ibu

X5 : Tingkat Kesejahteraan

Berdasarkan Tabel 7, didapatkan hasil bahwa variabel pendidikan ayah dan tingkat kesejahteraan tidak signifikan terhadap variabel keikutsertaan KB, hal ini ditunjukkan dengan nilai $W_j < \chi^2_{(1;0,05)} = 3,84$ atau nilai $p\text{-value} > \alpha = 0,05$. Karena ada dua variabel yang tidak signifikan maka dibentuk model baru yang akan diuji lagi dengan uji rasio likelihood ke-2 yang nilainya pada baris kedua dalam Tabel 6 dan uji wald kedua pada Tabel 8. Dari Tabel 8 didapatkan kesimpulan semua variabel yaitu umur ibu, jumlah anak dan pendidikan ibu signifikan yang ditunjukkan dengan $W_j > \chi^2_{(1;0,05)} = 3,84$ atau nilai $p\text{-value} > \alpha = 0,05$.

4.3.3. Uji Kesesuaian Model

Berdasarkan uji kesesuaian model diperoleh nilai $\chi^2 < \chi^2_{(0,05;7)}$ yaitu $7,495 < 14,067$ atau $p\text{-value} > \alpha$ yaitu $0,379 > 0,05$ sehingga H_0 diterima. Jadi, model regresi probit biner yang terbentuk sesuai atau tidak ada perbedaan antara observasi dengan kemungkinan hasil prediksi.

4.3.4. Model Akhir

Setelah dilakukan uji signifikansi terhadap model, baik secara keseluruhan maupun individual serta dilakukan uji kesesuaian model maka diperoleh model akhir sebagai berikut:

$$Y_i^* = 0,013 + 0,639x_1 - 1,067x_2 + 0,619x_{4(1)} + 0,215x_{4(2)}$$

4.4. Ketepatan Klasifikasi

Dari hasil perhitungan nilai Y_i^* setiap pengamatan dapat dibentuk tabel berikut:

Tabel 9. Matriks Konfusi Klasifikasi Metode Regresi Logistik Biner

Observasi	Prediksi		Akurasi (%)
	Tidak ikut KB	Ikut KB	
Tidak ikut KB	94	78	54,7
Ikut KB	46	175	79,2
Total			68,4

Berdasarkan Tabel 9, dapat dihitung nilai ketepatan klasifikasinya adalah 68,4% yang didapat dari membandingkan jumlah dari pasangan usia subur yang tepat diklasifikasikan dengan jumlah seluruh observasi dan nilai kesalahan klasifikasinya (APER) adalah 31,6%.

4.5. Perbandingan Metode Regresi Logistik Biner Dengan Metode Regresi Probit Biner

Tabel 10. Perbandingan Ketepatan Klasifikasi

Model	Ketepatan Klasifikasi (%)	
	APER	AKURASI
Regresi Logistik Biner	31,0	69,0
Regresi Probit Biner	31,6	68,4

Berdasarkan Tabel 10, nilai akurasi regresi logistik biner sebesar 69,0% dengan nilai APER 31,0% sedangkan nilai akurasi regresi probit biner sebesar 68,4% dengan nilai APER 31,6%. Hal ini menunjukkan bahwa pemodelan keikutsertaan KB di Kabupaten Semarang tahun 2014 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 kedua model sama baiknya untuk mengklasifikasikan keikutsertaan KB di Kabupaten Semarang.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan mengenai keikutsertaan KB di Kabupaten Semarang tahun 2014 menggunakan regresi logistik biner dan regresi probit biner, diperoleh kesimpulan sebagai berikut:

1. Keikutsertaan KB di Kabupaten Semarang tahun 2014 dipengaruhi oleh faktor umur ibu, jumlah anak yang dimiliki pasangan usia subur dan pendidikan ibu.
2. Nilai ketepatan klasifikasi metode regresi logistik biner adalah sebesar 69,0% dengan nilai kesalahan klasifikasi (APER) 31,0% dan nilai ketepatan klasifikasi metode regresi probit biner adalah sebesar 68,4% dengan nilai kesalahan klasifikasi (APER) 31,6%.
3. Dari hasil perbandingan ketepatan klasifikasi antara metode regresi logistik biner dan metode regresi probit biner, kedua metode relatif sama atau kedua metode tersebut sama baiknya untuk mengklasifikasikan keikutsertaan KB di Kabupaten Semarang tahun 2014.

DAFTAR PUSTAKA

- Agresti, A. 2002. *Categorical Data Analysis Second Edition*. John Willey and Sons. New York.
- BPS. 2010. *Hasil Sensus Penduduk 2010 Data Agregat Per Provinsi*. BPS. Jakarta.
- BKKBN. 2014. *Tata Cara Pelaksanaan Pencatatan Dan Pelaporan Program Kependudukan Dan Keluarga Berencana Nasional*. Semarang: BKKBN.
- Greene, W.H. 2000. *Econometrics Analysis Sixth Edition*. Prentice Hall: New Jersey.
- Hosmer, D.W., dan Lemeshow. 2000. *Applied Logistic Regression*. USA: John Willey and Sons.
- Irianto, K. 2014. *Pelayanan Keluarga Berencana*. Bandung: ALFABETA.
- Johnson, R. A. dan Wichern. D. W. 1992. *Applied Multivariate Statistical Analysis*. Prentice Hall: New Jersey.