

PERBANDINGAN METODE KLASIFIKASI REGRESI LOGISTIK BINER DAN NAIVE BAYES PADA STATUS PENGGUNA KB DI KOTA TEGAL TAHUN 2014

Nanci Rajagukguk¹, Dwi Ispriyanti², Yuciana Wilandari³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

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

ABSTRACT

Indonesia is a country that includes having the highest population density in the world. It is because the Indonesian state has a birth rate is so high. One of the efforts to control that population growth can be controlled by using the Keluarga Berencana program. In this study, the method used is the Binary Logistic Regression and Naive Bayes. To perform classification KB User Status in Tegal 2014, the variable used is the wife's age, the age of first marriage, type of wife's job, type of husband's job, wife's education, husband's education, and number of children. The training data comparison testing is 70:30. Based on the research results using binary logistic regression showed that a significant predictor variables that affect the status of keluarga Berencana user are wife's age, type of wife's job, and number of children with a classification accuracy of testing data 83.33%. While with the Naive Bayes method obtained classification accuracy of 81.75%. From this analysis it can be concluded that the Binary Logistic Regression method is better than the Naive Bayes in classifying the status of KB users in Tegal 2014.

Keywords : Binary Logistic Regression, Naive Bayes, Keluarga Berencana, Classification.

1. PENDAHULUAN

Indonesia merupakan negara yang termasuk memiliki jumlah penduduk terbanyak di dunia. Hal ini disebabkan karena negara Indonesia memiliki tingkat kelahiran yang begitu tinggi sehingga terjadilah kepadatan penduduk. Badan Kependudukan Keluarga Berencana Nasional (BKKBN) merupakan lembaga pemerintah non kementerian yang berada di bawah dan bertanggungjawab kepada Presiden melalui Menteri yang bertanggungjawab dibidang Kesehatan

Menurut Entjang(1986) dalam Garis-garis Besar Haluan Negara (GBHN) sebagai Ketetapan Majelis Permusyawaratan Rakyat (MPR) No.IV/MPR/1987 disebutkan bahwa program KB bertujuan untuk meningkatkan kesejahteraan ibu dan anak dalam rangka mewujudkan keluarga bahagia yang menjadi dasar bagi terwujudnya masyarakat yang sejahtera dengan mengendalikan kelahiran sekaligus dalam rangka menjamin terkendalinya pertumbuhan penduduk Indonesia. Ada banyak faktor yang mempengaruhi masyarakat dalam menggunakan KB. Dalam penelitian ini diasumsikan bahwa faktor-faktor yang mempengaruhi status menggunakan KB antara lain umur istri, umur pertama kawin istri, jenis pekerjaan istri, jenis pekerjaan suami, status pendidikan istri, status pendidikan suami dan jumlah anak. Pengguna KB diklasifikasikan dalam dua kategori yaitu pengguna dan tidak pengguna. Salah satu metode statistika untuk mengklasifikasikan adalah Regresi Logistik Biner dan Naive Bayes. Dalam penelitian ini akan dibandingkan kedua metode tersebut. Adapun tujuan dari penulisan tugas akhir ini adalah :

1. Memperoleh model dan ketepatan klasifikasi pengguna KB dengan menggunakan metode Regresi Logistik Biner
2. Memperoleh ketepatan klasifikasi dengan menggunakan metode Naive Bayes
3. Memperoleh perbandingan ketepatan klasifikasi Regresi Logistik Biner dan Naive Bayes pada data pengguna KB Kota Tegal tahun 2014.

2. TINJAUAN PUSTAKA

2.1 Kota Tegal

Kota Tegal adalah salah satu kota di Provinsi Jawa Tengah Indonesia yang berbatasan dengan Kabupaten Brebes di sebelah barat, Laut Jawa di sebelah utara, serta Kabupaten Tegal di sebelah timur dan selatan. Kota Tegal terdiri 4 kecamatan, yakni Tegal Barat, Tegal Timur, Tegal Selatan, dan Margadana.

2.2 Konsep Klasifikasi dan Pengukurannya

Menurut Prasetyo (2012), klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua proses yang dilakukan yaitu dengan membangun model untuk disimpan sebagai memori dan menggunakan model tersebut untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu data lain supaya diketahui di kelas mana objek data tersebut dimasukkan berdasarkan model yang telah disimpan dalam memori.

Sistem dalam klasifikasi diharapkan mampu melakukan klasifikasi semua set data dengan benar, namun tidak dapat dipungkiri bahwa kesalahan akan terjadi dalam proses pengklasifikasian tersebut sehingga perlunya dilakukan pengukuran kinerja dari sistem klasifikasi tersebut. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Contoh dari matriks konfusi untuk dua kelas (biner) dapat dilihat pada Tabel 1.

Tabel 1. Matriks Konfusi untuk Klasifikasi Dua kelas

f_{ij}		Kelas Hasil Prediksi (j)	
		kelas 1	kelas 0
Kelas Asli (i)	Kelas 1	f11	f10
	kelas 0	f01	f00

Akurasi hasil prediksi dapat dihitung ketika jumlah data yang diklasifikasi secara benar maupun salah telah diketahui. Untuk menghitung akurasi digunakan formula:

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{22}}{f_{11} + f_{12} + f_{21} + f_{22}}$$

Untuk menghitung kesalahan klasifikasi atau laju eror digunakan formula yaitu :

$$\text{Laju eror} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

2.3 Metode Regresi Logistik Biner

Model regresi logistik biner digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Menurut Hosmer dan Lemeshow (2000), bentuk model regresi logistik biner dengan variabel prediktor adalah sebagai berikut:

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}}$$

Sedangkan logit dari $\pi(x)$ adalah

$$g(x_i) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

2.3.1 Pengujian Model

Untuk menguji signifikansi dari parameter dalam model, digunakan uji rasio Likelihood dan uji Wald.

1. Uji rasio Likelihood

Uji rasio Likelihood digunakan untuk menguji kelayakan model yang diperoleh dari estimasi parameter yang bertujuan untuk mengetahui apakah variabel independen yang terdapat dalam model berpengaruh nyata atau tidak secara keseluruhan (Hosmer and Lemeshow, 2000)

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

H_1 : paling sedikit satu $\beta_j \neq 0, j = 1, 2, \dots, p$

Statistik uji : $G = -2 \ln \frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}}$

Kriteria uji : Tolak H_0 jika $G > \chi^2_{\alpha, p}$

2. Uji Wald

Uji Wald digunakan untuk mengetahui variabel-variabel bebas mana yang mempunyai hubungan lebih kuat dengan variabel respon nya (Hosmer and Lemeshow, 2000).

Hipotesa : $H_0 : \beta_j = 0$ dengan $j = 1, 2, \dots, p$

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

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

Kriteria uji : tolak H_0 jika $W_j > \chi^2_{\alpha, 1}$

3. Uji Kesesuaian Model

Uji kesesuaian model digunakan untuk menilai apakah model sesuai atau tidak.

Pada uji ini, statistik uji yang digunakan adalah uji Hosmer dan Lemeshow (Hosmer dan Lemeshow, 2000).

Hipotesis

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Statistik Uji : $\hat{C} = \sum_{f=1}^g \left[\frac{(O_f - n \bar{\pi}_f)^2}{(n_f \bar{\pi}_f (1 - \bar{\pi}_f))} \right]$

Dimana : g = jumlah group

O_f = jumlah nilai variabel respon pada group ke f

$\bar{\pi}_f$ = rata-rata taksiran peluang

n_f = banyak observasi pada grup ke- f

Kriteria Uji : H_0 ditolak jika $\hat{C} < \chi^2_{(g-2, \alpha)}$

2.4 Metode Naive Bayes

Naive Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian.

Menurut Prasetyo (2012), jika X merupakan vektor yang berisi fitur dan Y adalah label kelas, Naive Bayes dituliskan dengan $P(Y|X)$. Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y , sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*) Y .

Formulasi Naive Bayes untuk klasifikasi adalah sebagai berikut :

$$P(Y | X_j) = \frac{P(Y) \prod_{j=1}^p P(X_j | Y)}{P(X_j)}$$

Dimana:

- $P(Y|X_j)$ = probabilitas data dengan vektor X pada kelas Y .
- $P(Y)$ = probabilitas awal kelas Y .
- $\prod_{j=1}^p P(X_j | Y)$ = probabilitas independen kelas Y dari semua fitur dalam vektor X .
- Nilai $P(X_j)$ = probabilitas dari X_j .

Probabilitas $P(X)$ selalu tetap sehingga dalam perhitungan prediksi nantinya dapat dihilangkan dan hanya menghitung bagian $P(Y) \prod_{j=1}^p P(X_j | Y)$ saja dengan memilih nilai yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen $\prod_{j=1}^p P(X_j | Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y , yang dinotasikan dengan :

$$P(X | Y=y) = \prod_{j=1}^p P(X_j | Y = y)$$

Setiap set fitur $X = [X_1, X_2, X_3, \dots, X_p]$ terdiri atas p atribut

3. METODOLOGI

3.1 Jenis dan Sumber Data

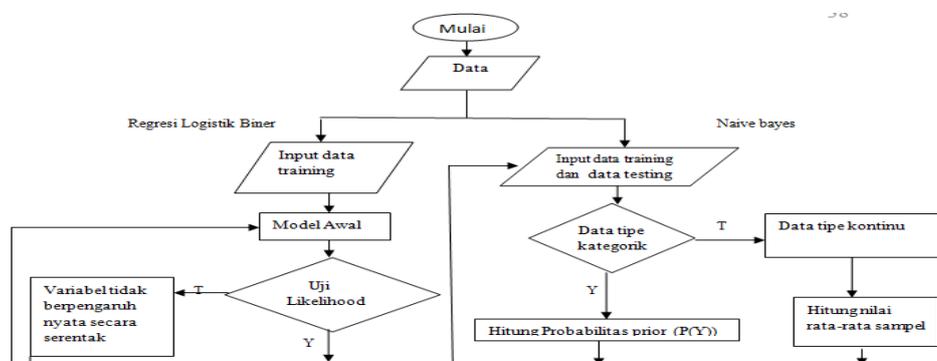
Data yang digunakan sebagai studi kasus pada tugas akhir ini berupa data sekunder yaitu data Pemutakhiran Data Keluarga (MDK) yang diambil dari hasil survei BKKBN (Badan Kependudukan Keluarga Berencana Nasional) Jawa Tengah untuk Kota Tegal pada Tahun 2014.

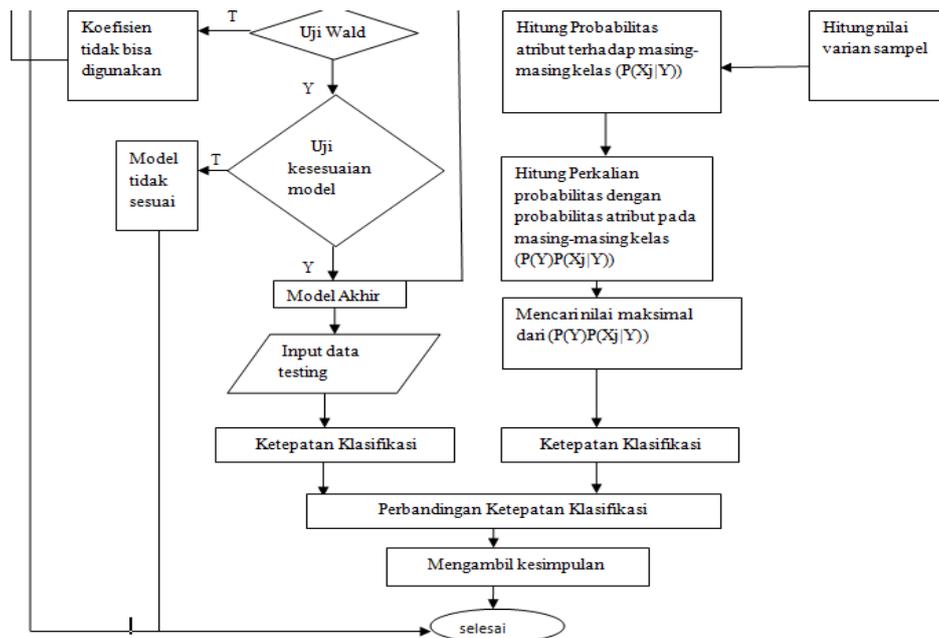
3.2 Variabel Data

Pada penelitian tugas akhir ini digunakan beberapa atribut atau variabel. Variabel yang akan digunakan dalam penelitian ini terbagi menjadi 2 yaitu variabel respon dan variabel prediktor.

- Variabel respon :
 Y = Status Pengguna KB, 1 = Pengguna KB 0 = bukan pengguna KB
- Variabel-Variabel prediktor :
 X_1 = umur istri, X_2 = usia kawin pertama istri, X_3 = jenis pekerjaan istri
 X_4 = Jenis pekerjaan suami, X_5 = status pendidikan istri, X_6 = status pendidikan suami, X_7 = jumlah anak

3.3. Teknik Pengolahan Data





4. ANALISIS DAN PEMBAHASAN

4.1 Analisis Status Pengguna KB dengan Metode Regresi Logistik Biner

Pada analisis data Status pengguna KB Kota Tegal tahun 2014, digunakan data training dan testing dengan perbandingan 70% : 30% dimana data *training* sebanyak 294 dan *testing* sebanyak 126. Variabel prediktor terdiri dari umur ibu, usia nikah pertama ibu, jenis pekerjaan istri, jenis pekerjaan suami, status pendidikan istri, status pendidikan suami, dan jumlah anak dan variabel respon yang digunakan ialah status pengguna KB (1 = Pengguna KB) dan (0 = Bukan Pengguna KB).

4.1.1 Model Awal

Untuk membentuk model regresi logistik biner terlebih dahulu dicari estimasi parameter yang diperoleh dari output pada lampiran 2. Hasilnya adalah sebagai berikut :

Tabel 3. Nilai Estimasi Parameter

Variabel	Estimasi	
Constant	2,898	
X1	-0,058	
X2	-0,038	
X3	X3(1)	-0,444
	X3(2)	-0,830
X4	X4(1)	-0,609
	X4(2)	0,039
X5	X5(1)	0,605
	X5(2)	0,826
X6	X6(1)	0,282
	X6(2)	-0,085
X7	0,597	

Berdasarkan estimasi parameter tersebut, diperoleh model sebagai berikut :

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

dengan nilai $g(x) = 2,898 - 0,058 X1 - 0,038 X2 - 0,444 X3(1) - 0,830 X3(2) - 0,609 X4(1) + 0,039 X4(2) + 0,605 X5(1) + 0,826 X5(2) + 0,282 X6(1) - 0,085 X6(2) + 0,597 X7$

4.1.2 Uji Rasio Likelihood

- Hipotesis
 - $H_0 : \beta_1 = \beta_2 \dots = \beta_{11} = 0$
 - $H_1 : \text{paling sedikit satu } \beta_j \neq 0, j = 1, 2, \dots, 11$
- Taraf signifikansi $\alpha = 5\%$
- Statistik Uji

$$G = -2 \ln \frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}}$$

$$= 33,70$$

- d. Kriteria Uji
Tolak H_0 jika $G > \chi^2_{\alpha,p}$ atau $\text{sig} < \alpha$
- e. Keputusan
Karena nilai statistik Uji yang diperoleh sebesar $33,70 > \chi^2_{0,05,11} = 19,68$ atau nilai $\text{sig } 0,000 < \alpha = 0,05$ maka H_0 ditolak.
- f. Kesimpulan
maka disimpulkan bahwa variabel prediktor yang terdapat pada model berpengaruh nyata secara serentak.

4.1.3 Uji Wald

- a. Hipotesis
 $H_0 : \beta_j = 0$ dengan $j = 1, 2, \dots, 11$
 $H_1 : \beta_j \neq 0$ dengan $j = 1, 2, \dots, 11$
- b. Taraf signifikansi $\alpha = 5\%$
- c. Statistik Uji
- $$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$
- d. Kriteria Uji
tolak H_0 jika $W_j > \chi^2_{\alpha,1}$ atau nilai $\text{sig} < \alpha$

Tabel 4. Hasil Uji Wald

Variabel	Wald	Sig	Keputusan
X1	6,064	0,014	Ho ditolak
X2	0,877	0,349	Ho diterima
X3(1)	0,171	0,679	Ho diterima
X3(2)	6,757	0,009	Ho ditolak
X4(1)	0,228	0,633	Ho diterima
X4(2)	0,001	0,974	Ho diterima
X5(1)	0,83	0,362	Ho diterima
X5(2)	1,894	0,169	Ho diterima
X6(1)	0,202	0,653	Ho diterima
X6(2)	0,023	0,878	Ho diterima
X7	9,542	0,002	Ho ditolak

Keterangan :

X1 = Umur istri

X2 = Umur menikah

X3 = Jenis pekerjaan istri

X4 = jenis pekerjaan suami

X5 = status pendidikan istri

X6 = status pendidikan suami

X7 = jumlah anak

- e. Keputusan
Berdasarkan pada Tabel 4 dapat diperoleh bahwa untuk variabel X2 X4, X5, dan X6 H_0 diterima, karena nilai uji Wald yaitu $W_j < \chi^2_{(0,05,1)} = 3,84$ atau nilai $\text{sig} > 0,05$ sedangkan untuk variabel X1, X3, dan X7 H_0 ditolak karena $W_j > \chi^2_{(0,05,1)} = 3,84$ atau nilai $\text{sig} < 0,05$
- f. Kesimpulan
Maka disimpulkan bahwa H_0 diterima yaitu variabel yang tidak berpengaruh signifikan terhadap variabel respon status pengguna KB ialah variabel X2, X4, X5, dan X6), sedangkan H_0 ditolak yaitu variabel X1, X3, dan X7, yang artinya berpengaruh signifikan.

Setelah dilakukan uji signifikansi terhadap model diperoleh hasil bahwa variabel yang berpengaruh signifikan terhadap status pengguna KB yaitu variabel X1 (umur istri), X3 (Jenis pekerjaan istri), dan X7 (jumlah anak), sehingga model Regresi Logistik Biner yang terbentuk adalah sebagai berikut :

$$\hat{\pi}(x) = \frac{e^{2,898 - 0,058 X1 - 0,444 X3(1) - 0,830 X3(2) + 0,597 X7}}{1 + e^{2,898 - 0,058 X1 - 0,830 X3(2) + 0,597 X7}}$$

Selanjutnya dari ketiga variabel yang berpengaruh signifikan tersebut, dilakukan analisis kembali untuk memperoleh model akhir yang sesuai. Berikut analisis model Regresi Logistik biner dengan tidak mengituksertakan variabel yang tidak berpengaruh signifikan.

4.1.4 Uji Rasio Likelihood II

- Hipotesis
 $H_0 : \beta_1 = \beta_2 \dots = \beta_4 = 0$
 $H_1 : \text{paling sedikit satu } \beta_j \neq 0, j = 1, 2, 3, 4$
- Taraf signifikansi $\alpha = 5\%$
- Statistik Uji

$$G = -2 \ln \frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}}$$

$$= 26,14$$
- Kriteria Uji
Tolak H_0 jika $G > \chi^2_{(0,05,4)}$ atau $\text{sig} < \alpha$
- Keputusan
Karena nilai statistik Uji yang diperoleh sebesar $26,14 > \chi^2_{0,05,4} = 9,49$ atau nilai $\text{sig } 0,000 < \alpha = 0,05$ maka H_0 ditolak
- Kesimpulan
Maka disimpulkan bahwa variabel independen yang terdapat pada model berpengaruh nyata secara serentak.

4.1.5 Uji Wald II

- Hipotesis
 $H_0 : \beta_j = 0$ dengan $j = 1, 2, 3, 4$
 $H_1 : \beta_j \neq 0$ dengan $j = 1, 2, 3, 4$
- Taraf signifikansi $\alpha = 5\%$
- Statistik Uji :

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$
- Kriteria Uji
Tolak H_0 jika $W_j > \chi^2_{\alpha,1}$ atau nilai $\text{sig} < \alpha$

Tabel 5. Hasil Uji Wald II

Variabel	Wald	sig	Keputusan
X1	6,713	0,00	Ho ditolak
X3(1)	1,651	0,199	Ho diterima
X3(2)	7,174	0,007	Ho ditolak
X7	11,711	0,001	Ho ditolak

- Keputusan
Berdasarkan pada Tabel 5 dapat diperoleh bahwa variabel X1, X3, dan X7, Ho ditolak karena $W_j > \chi^2_{(0,05,1)} = 3,84$ atau nilai $\text{sig} < 0,05$

- f. Kesimpulan
Berdasarkan keputusan H_0 ditolak disimpulkan bahwa variabel X_1 , X_3 , dan X_7 , berpengaruh signifikan terhadap status pengguna KB.

4.1.6 Uji kesesuaian Model

a. Hipotesis

H_0 : Model sesuai

H_1 : Model tidak sesuai

b. Taraf signifikansi $\alpha = 5\%$

c. Statistik Uji

$$\hat{C} = \sum_{f=1}^g \left[\frac{(O_f - n \bar{\pi}_f)^2}{(n_f \bar{\pi}_f (1 - \pi_f))} \right]$$

Nilai $C = 6,575$ dan sig $0,583$

d. Kriteria Uji

Tolak H_0 jika $C > \chi^2_{(0,05; 8)}$ atau sig $< \alpha$

e. Keputusan

Nilai sig = $0,583 > \alpha = 0,05$ maka H_0 diterima

f. Keputusan

Berdasarkan keputusan H_0 diterima maka dapat disimpulkan bahwa model Sesuai.

4.1.7 Model Akhir

Setelah dilakukan uji signifikansi terhadap model, maka diperoleh model akhir untuk data training sebagai berikut :

$$\hat{\pi}(x) = \frac{e^{2,653 - 0,058X_1 - 1,261X_3(1) - 0,813X_3(2) + 0,650X_7}}{1 + e^{2,653 - 0,058X_1 - 1,261X_3(1) - 0,813X_3(2) + 0,650X_7}}$$

4.1.8 Estimasi dan Interpretasi Model Regresi Logistik Biner

A. Odds Ratio

Tabel 6. Nilai Odds Ratio

Variabel	B	Exp(B)
X1	-0,058	0,944
X3(1)	-1,261	0,283
X3(2)	-0,813	0,443
X7	0,65	1,915

Berdasarkan Tabel 6 diperoleh informasi sebagai berikut :

- Variabel X_1 , Nilai Odds ratio sebesar 0,944 artinya bahwa setiap bertambahnya umur istri resiko kecenderungan seorang istri untuk menjadi pengguna KB turun sebesar 0,056.
- Variabel $X_3(1)$, Nilai odds ratio sebesar 0,283 artinya bahwa resiko kecenderungan seorang istri yang bekerja di pemerintahan untuk menjadi pengguna KB adalah 0,283 kalinya dari istri yang tidak bekerja.
- Variabel $X_3(2)$, nilai Odds ratio sebesar 0,443 artinya bahwa kecenderungan seorang istri yang bekerja di non pemerintahan untuk menjadi pengguna KB adalah 0,443 kalinya dari istri yang tidak bekerja.
- Variabel X_7 , Nilai Odds ratio sebesar 1,915 artinya bahwa setiap bertambahnya jumlah anak resiko kecenderungan seorang istri untuk menjadi pengguna KB naik sebesar 0,915.

B. Interpretasi Model Terbaik

- Peluang seorang istri dengan umur 30 tahun, pekerjaan non pemerintah, dan jumlah anak dua untuk menjadi pengguna KB adalah :

$$\hat{\pi}(x) = \frac{e^{2,653-0,058(30)-1,261(0)-0,813(1)+0,650(2)}}{1+e^{2,653-0,058(30)-1,261(0)-0,813(1)+0,650(2)}} = 0,8021$$

- Peluang seorang istri dengan umur 48 tahun, pekerjaan non pemerintah, dan jumlah anak satu untuk menjadi bukan pengguna KB adalah :

$$1 - \hat{\pi}(x) = \frac{1}{1+e^{2,653-0,058(48)-1,261(0)-0,813(1)+0,650(1)}} = 0,573$$

- Peluang seorang istri dengan umur 41 tahun, pekerjaan non pemerintah, dan jumlah anak empat untuk menjadi bukan pengguna KB adalah :

$$1 - \hat{\pi}(x) = \frac{1}{1+e^{2,653-0,058(41)-1,261(0)-0,813(1)+0,650(4)}} = 0,113$$

4.2 Ketepatan klasifikasi Menggunakan Regresi Logistik Biner

Setelah dilakukan uji signifikansi parameter maka akan dilakukan perhitungan ketepatan klasifikasi terhadap model. Tabel dibawah ini adalah hasil ketepatan klasifikasi.

Tabel 7. Matriks Konfusi Regresi Logistik Biner

<i>fij</i>		kelas Hasil Prediksi		
		kelas 1	kelas 0	
Kelas Asli	kelas 1	104	1	83,33%
	kelas0	20	1	

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{105}{126} = 0,83333$$

$$\text{Laju Error} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{21}{126} = 0,16667$$

Berdasarkan Tabel 7 dan perhitungan diperoleh informasi bahwa hasil ketepatan klasifikasi menggunakan Regresi Logistik Biner menghasilkan ketepatan Klasifikasi sebesar 0,83333 atau 83,33% sedangkan laju error nya sebesar 0,16667 atau 16,67%.

4.3 Pengklasifikasian dengan Metode Naive Bayes

Pengolahan data dengan metode Naive Bayes menggunakan variabel-variabel yang signifikan yang diperoleh dari model akhir pada pembahasan Regresi Logistik Biner yaitu Umur istri, Pekerjaan istri, dan Jumlah Anak. Proporsi data *training* dan *testing* yang digunakan yaitu 70% : 30% dimana data *training* sebanyak 294 dan *testing* sebanyak 126 data dan menghasilkan ketepatan klasifikasi sebesar 81,75% dan laju error sebesar 0.1825 atau 18,25%.

Berdasarkan Tabel 8 diperoleh informasi bahwa yang tepat dikategorikan sebagai pengguna KB yaitu sebanyak 103 orang. Sedangkan yang dikategorikan tepat sebagai bukan pengguna KB yaitu sebanyak 2 orang.

Tabel 8 . Matriks Naive Bayes

<i>fij</i>		kelas Hasil Prediksi		
		kelas 1	kelas 0	
Kelas Asli	kelas 1	103	2	81,75%
	kelas0	21	0	

Dari Tabel 8 dapat dihitung nilai ketepatan klasifikasi sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{103}{126} = 0,8175$$

$$\text{Laju Error} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{23}{126} = 0,1825$$

4.4 Perbandingan Ketepatan Klasifikasi

Tabel 9. Perbandingan Ketepatan Klasifikasi

Training : Testing	Regresi Logistik Biner	Naive bayes
95% : 5%	71,42%	71,43%
90% : 10%	80,95%	78,57%
80% : 20%	72,61%	72,62%
70% : 30%	83,33%	81,75%

Dari hasil pengklasifikasian pada Tabel 9 dapat disimpulkan bahwa metode Regresi Logistik Biner bekerja lebih baik dibandingkan dengan Naive Bayes dalam mengklasifikasikan status pengguna KB.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan pada data *training* dan *testing* yaitu dengan perbandingan 70% dan 30% dimana data *training* sebanyak 294 dan *testing* sebanyak 126, maka dapat diambil kesimpulan sebagai berikut:

1. Model terbaik dari variabel-variabel yang signifikan pada metode Regresi Logistik Biner adalah

$$\hat{\pi}(x) = \frac{e^{2,653 - 0,058X_1 - 1,261X_3(1) - 0,813X_3(2) + 0,650X_7}}{1 + e^{2,653 - 0,058X_1 - 1,261X_3(1) - 0,813X_3(2) + 0,650X_7}}$$

2. Ketepatan klasifikasi dengan metode Regresi Logistik Biner sebesar 83,33 % dengan laju error sebesar 16,67% sedangkan ketepatan klasifikasi dengan metode Naive Bayes sebesar 81,75 % dengan laju error sebesar 18,25%. Pengklasifikasian menggunakan metode Regresi Logistik Biner lebih baik dibandingkan dengan metode Naive Bayes.

6. DAFTAR PUSTAKA

- Agresti, A. 2002. *Categorical Data Analysis Second Edition*. Jhon Wiley & Sons, Inc: USA
- Bain, L.J, and Engelhardt, M, 1991. *Introduction to Probability and Mathematical Statistic*. Second Edition. Duxbury Press:California
- Entjang, I.1986. *Pendidikan Kependudukan dan Keluarga Berencana*. Penerbit Alumni. Bandung
- Hartanto, H.1994. *Keluarga Berencana dan Kontrasepsi*. Jakarta: Pustaka Sinar Harapan
- Hosmer, D.W. and Lemeshow, S.2000. *Applied Logistic Regression Second Edition*. Jhon Wiley & Sons, Inc: USA
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI Yogyakarta
- Subanar. 2006. *Pengantar Teori Ukuran dan Probabilitas*. Yogyakarta: FMIPA UGM www.tegalkota.go.id (diakses pada 5 September 2014)