

KLASIFIKASI REGRESI LOGISTIK MULTINOMIAL DAN FUZZY K-NEAREST NEIGHBOR (FK-NN) DALAM PEMILIHAN METODE KONTRASEPSI DI KECAMATAN BULAKAMBA, KABUPATEN BREBES, JAWA TENGAH

Erysta Risky Rismia^{1*}, Tatik Widiharih², Rukun Santoso³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*Email : erystarr@gmail.com

ABSTRACT

The characteristics of society in choosing contraceptive methods are also the crucial factors for the government to prepare the family planning services needed at Bulakamba District, Brebes Regency, Central Java. In this case, a classification process needs to be done to assist the process of classifying the characteristics of society in the selection of contraceptive methods. Multinomial Logistic Regression classification is good in exploring data information meanwhile Fuzzy K Nearest Neighbor (FK-NN) classification is good for handling big data and noise. These two methods used in this study because they are relevant to the data applied and will be compared their classification accuracy through APER and Press's Q calculations. The classification accuracy results obtained based on the APER calculation for Multinomial Logistic Regression is 88,25% and Fuzzy K Nearest Neighbor (FK-NN) is 88,92%. Meanwhile, the Press's Q value of both methods are 9600,945 and 9518,014 greater than $\chi^2_{\alpha,1}$ which is 3,841, so that it is statistically accurate. Based on the results obtained, it can be concluded that Multinomial Logistic Regression classification method has a better classification accuracy than Fuzzy K Nearest Neighbor (FK-NN) method.

Keywords: *Contraceptive Methods, Multinomial Logistic Regression, FKNN, APER, Press's Q.*

1. PENDAHULUAN

Program Keluarga Berencana merupakan program skala nasional yang dikelola oleh Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) untuk mengendalikan pertumbuhan jumlah penduduk dan menekan angka kelahiran di Indonesia. Sejalan dengan terus berkembangannya Program KB yang dilaksanakan, pemerintah juga masih terus meningkatkan kualitas pelayanan KB untuk menunjang daya tarik masyarakat pada Program Keluarga Berencana tersebut salah satunya dengan menjamin ketersediaan alat/metode kontrasepsi. Karakteristik masyarakat dalam pemilihan metode kontrasepsi merupakan faktor penting bagi pemerintah untuk mempersiapkan pelayanan KB yang dibutuhkan di suatu daerah. Sehingga hal ini memunculkan peluang untuk dilakukannya proses klasifikasi guna membentuk ukuran-ukuran yang baku dalam pengelompokan dan pendefinisian karakteristik masyarakat dalam pemilihan metode kontrasepsi.

Klasifikasi merupakan metode penelitian ketika suatu objek data dinilai untuk dimasukkan ke dalam suatu kelas tertentu dari sejumlah kelas yang ada atau tersedia (Prasetyo, 2012). Terdapat beberapa metode klasifikasi yang dapat digunakan dalam proses klasifikasi, diantaranya Regresi Logistik Multinomial dan *Fuzzy K Nearest Neighbor* (FK-NN). Regresi Logistik Multinomial merupakan Regresi Logistik yang digunakan ketika variabel responnya memiliki skala yang bersifat *polychotomus* (Hosmer dan Lemeshow, 2000). Sedangkan *Fuzzy K Nearest Neighbor* (FK-NN) merupakan hasil perbaikan dari metode K Nearest Neighbor (K-NN). Perbaikan yang dilakukan oleh Fuzzy K-Nearest Neighbor (FK-NN) yaitu dengan menggabungkan konsep tetangga terdekat dan teorema Fuzzy dalam menyampaikan definisi pemberian label kelas pada kelas data uji yang diprediksi (Prasetyo, 2012).

Beberapa penelitian sebelumnya yang dijadikan referensi dalam penelitian ini yaitu penelitian dari Kristiani dkk (2015), yang mengaplikasikan *Multivariate Adaptive Regression Spline* (MARS) dan *Fuzzy K-Nearest Neighbor* (FK-NN) pada data kelompok

rumah tangga di Kabupaten Blora. Kesimpulan yang memperoleh yaitu ketepatan klasifikasi *Fuzzy K-Nearest Neighbor* (FK-NN) lebih baik dibandingkan metode *Multivariate Adaptive Regression Spline* (MARS). Selanjutnya penelitian oleh Puspita dkk (2014) dengan mengaplikasikan metode Regresi Logistik Ordinal dan *Fuzzy K-Nearest Neighbor* pada Tingkat Keluarga Sejahtera yang menghasilkan kesimpulan bahwa ketepatan klasifikasi dengan menggunakan *Fuzzy K-Nearest Neighbor* (FK-NN) lebih baik daripada klasifikasi Regresi Logistik Ordinal. Dari referensi tersebut, dalam penelitian ini dilakukan analisis Regresi Logistik Multinomial dan *Fuzzy K-Nearest Neighbor* (FK-NN) jika diaplikasikan pada kasus data pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah. Kedua metode klasifikasi tersebut dibandingkan ketepatan klasifikasinya melalui perhitungan APER dan Press's Q untuk mengetahui metode klasifikasi yang memiliki performa lebih baik dalam pengelompokan dan pendefinisian karakteristik pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1. Metode Kontrasepsi

Metode kontrasepsi memiliki peranan penting dalam meningkatkan pelayanan dan daya tarik masyarakat terhadap Program Keluarga Berencana. Menurut Peraturan No. 24 Tahun 2017 metode kontrasepsi secara umum dikelompokkan menjadi 8 jenis, yaitu metode operasi wanita (MOW), metode operasi pria (MOP), IUD (*Intrauterine Device*), implan, suntik, pil, kondom, dan tradisional.

2.2. Regresi Logistik Multinomial

Analisis Regresi Logistik Multinomial merupakan Regresi Logistik yang digunakan ketika variabel respon mempunyai skala yang bersifat *polychotomus* atau multinomial (Hosmer dan Lemeshow, 2000). Sedangkan variabel prediktor pada analisis Regresi Logistik Multinomial bisa berupa data kategorik maupun numerik.

2.2.1. Pembentukan Model

Secara umum model Regresi Logistik Multinomial dengan p merupakan variabel prediktor dapat dituliskan sebagai berikut :

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

Dengan $\pi(x)$ merupakan peluang kejadian dari y jika diberikan nilai x, dan y merupakan variabel respon yang bersifat kategorik serta x merupakan variabel prediktor. Untuk mempermudah pendugaan parameter maka digunakan transformasi logit dari $\pi(x)$ yang kemudian menghasilkan persamaan baru sebagai berikut :

$$g(x) = \ln \left\{ \frac{\pi(x)}{1 - \pi(x)} \right\} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2)$$

Dari persamaan (2), model Regresi Logistik pada persamaan (1) dapat dituliskan secara lebih sederhana seperti berikut :

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (3)$$

2.2.2. Estimasi Parameter

Dalam menentukan estimasi parameter dari Regresi Logistik Multinomial digunakan metode Maksimum Likelihood yang selanjutnya akan diselesaikan dengan metode Iterasi Newton Raphson. Misalkan sebuah variabel respon memiliki 3 kategori yaitu $j = 0, 1, 2$, maka fungsi likelihood untuk sebuah sampel dari n pengamatan secara independen adalah sebagai berikut :

$$l(\beta) = \prod_{i=1}^n \pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \quad (4)$$

Dari persamaan (4) kemudian dibentuk fungsi ln-likelihood sebagai berikut :

$$L(\beta) = \sum_{i=1}^n y_{1i}g_1(x_i) + y_{2i}g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \quad (5)$$

Dari persamaan (5) kemudian didapatkan bentuk umum turunan pertamanya sebagai berikut:

$$\frac{\partial L(\beta)}{\partial \beta_{j_0}} = \sum_{i=1}^n [y_{ji} - \pi_j(x_i)] \quad (6) \quad \frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} [y_{ji} - \pi_j(x_i)] \quad (7)$$

Hasil turunan pertama dapat dinyatakan dalam bentuk matrik $\mathbf{M}(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}))$, dengan penjabaran (untuk 3 kategori) :

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ x_{11} & x_{11} & \dots & x_{1n} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ x_{p1} & x_{p1} & \dots & x_{pn} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} \quad (\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x})) = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix} - \begin{pmatrix} \pi_{11} \\ \pi_{12} \\ \vdots \\ \pi_{2n} \\ \pi_{21} \\ \pi_{22} \\ \vdots \\ \pi_{2n} \end{pmatrix}$$

Hasil turunan pertama tidak berbentuk *close form*, maka dilanjutkan dengan iterasi Newton Raphson. Bentuk umum turunan kedua dapat dituliskan melalui persamaan berikut :

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{jk}} = - \sum_{i=1}^n x_{ki} x_{ki} \pi_j(x_i) [1 - \pi_j(x_i)] \quad (8)$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{jk}} = - \sum_{i=1}^n x_{ki} x_{ki} \pi_j(x_i) \pi_j(x_i) \quad (9)$$

Misalkan $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})$ adalah matrik yang berukuran $p \times p$ dari persamaan (8) dan (9) yang bernilai negatif.

$$\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{11} & \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{12} \\ \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{21} & \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{22} \end{bmatrix} \quad (10)$$

Dengan : $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{11} = \mathbf{X}'\mathbf{V}_1\mathbf{X}$ $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{12} = \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{21} = \mathbf{X}'\mathbf{V}_3\mathbf{X}$ $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})_{22} = \mathbf{X}'\mathbf{V}_2\mathbf{X}$
 dan \mathbf{V}_j merupakan matrik diagonal $n \times n$ dengan elemen umum $\pi_j(x_i)[1 - \pi_j(x_i)]$, untuk $j = 1, 2$ dan $i = 1, 2, 3, \dots, n$. Untuk \mathbf{V}_3 merupakan matrik diagonal $n \times n$ dengan elemen umum $\pi_1(x_i)\pi_2(x_i)$. Selanjutnya akan diselesaikan dengan metode Iterasi Newton Raphson dengan langkah langkah sebagai berikut :

1. Memilih taksiran awal untuk $\boldsymbol{\beta}$, dimisalkan $\hat{\boldsymbol{\beta}} = 0$
2. Menghitung $\mathbf{M}(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}))$ dan $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})$, selanjutnya dihitung invers dari $\hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})$
3. Pada setiap $t+1$ dihitung taksiran baru yaitu

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \hat{\mathbf{i}}(\hat{\boldsymbol{\beta}})^{-1} \{ \mathbf{M}(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x})) \} \quad (11)$$
4. Iterasi berakhir jika diperoleh $|\hat{\boldsymbol{\beta}}^{(t+1)}| \cong |\hat{\boldsymbol{\beta}}^{(t)}|$

2.2.3. Pengujian Parameter

1) Uji Parameter secara Keseluruhan

Uji Rasio Likelihood merupakan uji signifikansi parameter secara keseluruhan atau bersama-sama (Hosmer dan Lemeshow, 2000).

Hipotesis :

$H_0: \beta_{jk} = 0$ (Tidak ada pengaruh antara variable prediktor terhadap variabel respon dengan $j = 1, 2, \dots, r-1$ dan $k = 1, 2, \dots, p$)

$H_1: \text{Paling sedikit ada satu } \beta_{jk} \neq 0$, dengan $j = 1, 2, \dots, r-1$ dan $k = 1, 2, \dots, p$
 (Paling sedikit ada satu variabel prediktor yang berpengaruh terhadap variabel respon)

Statistik uji: $G^2 = -2 \ln \left[\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right] \quad (12)$

Kriteria penolakan : Tolak H_0 apabila $G^2 > \chi^2_{(\alpha, db)}$ dengan derajat bebas yang digunakan yaitu sebesar banyaknya variabel predictor.

2) Uji Parameter secara Parsial

Bewick dkk (2005) mengatakan bahwa uji Wald digunakan untuk menguji signifikansi koefisien parameter secara individual dalam suatu model.

Hipotesis :

$H_0: \beta_{jk} = 0$, (Tidak ada pengaruh variabel prediktor ke- jk terhadap variabel respon, dengan $j = 1, 2, \dots, r-1$ dan $k = 1, 2, \dots, p$)

$H_1: \beta_{jk} \neq 0$, (Ada pengaruh variabel prediktor ke- jk terhadap variabel respon, dengan $j = 1, 2, \dots, r-1$ dan $k = 1, 2, \dots, p$)

$$\text{Statistik uji : } W_{jk} = \left\{ \frac{\hat{\beta}_{jk}}{SE(\hat{\beta}_{jk})} \right\}^2 \quad (13) \quad SE(\hat{\beta}_{jk}) = \sqrt{(\sigma^2(\hat{\beta}_{jk}))} \quad (14)$$

Dengan $SE(\hat{\beta}_{jk})$ merupakan dugaan kesalahan baku dari koefisien β_{jk} dan $\hat{\beta}_{jk}$.

Kriteria penolakan : Tolak H_0 apabila $W_{jk} > \chi^2_{(\alpha;1)}$.

2.2.4. Uji Kesesuaian Model

Uji kesesuaian model digunakan untuk menguji kelayakan suatu model yang telah terbentuk.

Hipotesis :

H_0 : Model sesuai (Tidak ada perbedaan antara hasil pengamatan dengan hasil prediksi)

H_1 : Model tidak sesuai (Ada perbedaan antara hasil pengamatan dengan hasil prediksi)

$$\text{Statistik uji : } \hat{C} = \sum_{i=1}^k \frac{(O_i - n_i \bar{\pi}_i)^2}{n_i \bar{\pi}_i (1 - \bar{\pi}_i)} \quad (15)$$

Dengan keterangan sebagai berikut :

O_i : Pengamatan pada kelompok ke- i

k : Jumlah kelompok (kombinasi kategori) dalam model serentak

n_i : Banyaknya pengamatan pada kelompok ke- i

$\bar{\pi}_i$: Rata-rata taksiran peluang pengamatan kelompok ke- i

Kriteria penolakan : Tolak H_0 apabila nilai $\hat{C} > \chi^2_{(\alpha;db)}$, dengan derajat bebas yang digunakan yaitu $db = (k-2)$.

2.2.5. Proses Klasifikasi

Dalam penentuan hasil prediksi dari suatu data uji maka harus dilihat besar estimasi probabilitas dari data tersebut pada setiap kategorinya. Kategori yang memiliki nilai estimasi probabilitas tertinggi akan dipilih sebagai kelas prediksi dari suatu data uji.

2.3. Fuzzy K-Nearest Neighbor (FK-NN)

Dalam algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) pengukuran nilai jarak antara data latih dan data uji pada data rasio/interval dapat digunakan rumus jarak *Euclidean* yang diformulasikan sebagai berikut (Li dkk, 2007) :

$$d(x_i, x_h) = \left(\sum_{l=1}^L |x_{il} - x_{hl}|^2 \right)^{\frac{1}{2}} \quad (16)$$

Dengan : $d(x_i, x_h)$ = jarak antara data x_i ke data x_h ; L = banyaknya variabel prediktor

Sedangkan pada data bersifat kategorik nominal, perhitungan jaraknya dapat menggunakan formula berikut :

$$d_{nom} = \begin{cases} 0 & \text{jika } x_{il} = x_{hl} \\ 1 & \text{jika } x_{il} \neq x_{hl} \end{cases} \quad (17)$$

Dengan : d_{nom} = jarak untuk kasus data nominal

x_{il} = data latih ke- i pada variabel prediktor ke- l

x_{hl} = data uji ke- h pada variabel prediktor ke- l

Pengukuran jarak untuk kasus data variabel kategorik bersifat ordinal yaitu dengan menggunakan formula berikut:

$$(x_{fl})_{baru} = \frac{R_{fl}-1}{M_l-1} \quad (18)$$

Dengan : $(x_{fl})_{baru}$ = data ke-f hasil tranformasi pada variabel prediktor ke-l.

R_{fl} = rangking data ke-f pada variabel ke-l.

M_l = total kategori pada variabel ke – l.

Nilai jarak *Euclidean* antara data latih dan data uji diurutkan dari yang terkecil dan diambil sebanyak K jarak terkecil untuk kemudian menjadi data tetangga yang disebut x_k , dimana $k = 1, 2, \dots, K$. Nilai keanggotaan data tetangga x_k ke kelas c_j dapat diformulasikan sebagai berikut :

$$u(x_k, c_j) = \begin{cases} 0 & \text{jika } x_k \text{ bukan milik kelas } c_j \\ 1 & \text{jika } x_k \text{ milik kelas } c_j \end{cases} \quad (19)$$

Menurut Li dkk (2007), formula untuk mencari nilai keanggotaan data uji x_i kedalam kelas c_j dapat dituliskan sebagai berikut :

$$u(x_i, c_j) = \frac{\sum_{k=1}^K [u(x_k, c_j)] \times [d(x_i, x_k)]^{\frac{-2}{(m-1)}}}{\sum_{k=1}^K d(x_i, x_k)]^{\frac{-2}{(m-1)}}} \quad (20)$$

Dengan keterangan sebagai berikut :

c_j : kelas/kelompok ke-j, dengan $1 \leq j \leq C$

C : banyaknya kelas/kelompok

x_i : data uji ke-i

x_k : data tetangga ke-k dalam K tetangga terdekat

$u(x_i, c_j)$: nilai keanggotaan data uji x_i kedalam kelas c_j .

K : banyaknya tetangga terdekat yang digunakan

$u(x_k, c_j)$: nilai keanggotaan dari data tetangga x_k dalam K tetangga terdekat pada kelas c_j yang dapat dicari dengan menggunakan persamaan (19).

$d(x_i, x_k)$: jarak *Euclidean* dari data x_i terhadap data x_k dalam K tetangga terdekat.

m : pembobot pangkat, dengan besar $m > 1$. Pembobot m ditentukan sesuai dengan banyaknya kelas yang terbentuk dalam klasifikasi atau kategori variabel dependen yang telah ditetapkan.

Konsep kerja algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) yaitu sebagai berikut :

1.) Menentukan K tetangga terdekat dengan perhitungan jarak *Euclidean* pada persamaan (16). K yang digunakan dalam penelitian ini yaitu $K = 1, 2, \dots, 15$.

2.) Menghitung nilai keanggotaan data uji pada setiap kelas dengan persamaan (20).

3.) Kelas dengan nilai keanggotaan data terbesar akan terpilih menjadi label kelas pada data uji, atau dapat dituliskan dengan formula berikut (Prasetyo, 2012):

$$y' = \text{ArgMax}_{j=1}^C (u(x_i, c_j)) \quad (21)$$

2.4. Ketepatan Klasifikasi

Nilai APER berfungsi untuk menghitung nilai proporsi sampel yang salah diklasifikasikan dalam proses klasifikasi (Johnson dan Winchern, 2007). Nilai APER dapat dihitung dengan menggunakan rumus berikut :

$$APER = \frac{\text{proporsi pengamatan yang terklasifikasikan salah}}{\text{total pengamatan}} \quad (22)$$

$$\text{Ketepatan klasifikasi} = 1 - APER \quad (23)$$

Semakin tinggi nilai ketepatan klasifikasi yang dihasilkan maka semakin baik pula keefektifan metode klasifikasi tersebut dalam mengelompokan suatu objek. Sedangkan Statistik uji dari *Press's Q* dapat dirumuskan sebagai berikut :

$$Press's Q = \frac{[N-(rC)]^2}{N(C-1)} \quad (24)$$

Dengan : C= banyaknya kelas/kelompok

N= banyaknya pengamatan secara keseluruhan

r= banyaknya pengamatan yang terklasifikasikan secara tepat

Proses klasifikasi akan dikatakan akurat secara statistik apabila nilai *Press's Q* lebih besar dari pada nilai kritis yang diambil dari tabel *Chi-Square* dengan derajat bebas bernilai 1 dan tingkat keyakinan sebesar α ($Press's Q > \chi_{\alpha,1}^2$).

3. METODE PENELITIAN

3.1. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data pemilihan metode kontrasepsi oleh PUS yang sedang menjalankan Program KB di Kabupaten Brebes, Kecamatan Bulakamba yang dapat diakses melalui website resmi Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Jawa Tengah dengan alamat : <http://pk.bkkbn.go.id/pk/Peralatan/UnduhFormulirDataKeluarga.aspx>.

3.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini yaitu sebagai berikut :

Tabel 1. Variabel Penelitian

Nama Variabel	Kategori
Variabel Respon (Y)	
Metode Kontrasepsi (Y)	1= IUD ; 2= MOP ; 3= Suntik ; 4= Kondom; 5= MOW ; 6= Implant ; 7= Pil ; 8= Tradisional.
Variabel Prediktor (X)	
Usia Suami (X1)	1= Umur < 20 tahun ; 2= 20 ≤ Umur ≤ 35 tahun ; 3= Umur > 35 tahun
Usia Istri (X2)	1= Umur < 20 tahun ; 2= 20 ≤ Umur ≤ 35 tahun ; 3= Umur > 35 tahun
Usia Kawin Pertama Suami (X3)	1 = Umur ≤ 25 tahun ; 2= Umur > 25 tahun
Usia Kawin Pertama Istri (X4)	1 = Umur ≤ 20 tahun ; 2= Umur > 20 tahun
Jumlah Anak Masih Hidup (X5)	1 = Jumlah anak hidup ≤ 2 anak ; 2 = Jumlah anak hidup > 2 anak
Pendidikan Suami (X6)	1 = Tidak sekolah ; 2 = Tamat SD/MI ; 3 = Tamat SMP/MTs ; 4 = Tamat SMA/Mta ; 5 = Tamat PT/Akademi
Pekerjaan Suami (X7)	1 = Tidak bekerja ; 2 = Kelompok pekerjaan tidak terampil ; 3 = Kelompok pekerjaan terampil
Pendidikan Istri (X8)	1 = Tidak sekolah ; 2 = Tamat SD/MI ; 3 = Tamat SMP/MTs ; 4 = Tamat SMA/Mta ; 5 = Tamat PT/Akademi
Pekerjaan Istri (X9)	1 = Tidak bekerja ; 2 = Kelompok pekerjaan tidak terampil ; 3 = Kelompok pekerjaan terampil
JKN Suami (X10)	1 = Tidak Memiliki ; 2 = Memiliki
JKN Istri (X11)	1 = Tidak Memiliki ; 2 = Memiliki
Tempat Pelayanan (X12)	1 = Lainnya (Penanganan pribadi) ; 2 = Fasilitas kesehatan tingkat pertama / FKTP ; 3 = Fasilitas kesehatan tingkat lanjut / FKTL

Ingin Punya Anak Lagi (X13)	1 = Tidak ingin memiliki anak lagi ; 2 = Ingin memiliki anak lagi
-----------------------------	---

3.3. Tahapan Pengolahan Data

Software yang digunakan dalam pengolahan penelitian ini adalah Program R versi 4.0.2 dan MATLAB 2015a. Tahapan analisis pada penelitian Tugas Akhir ini dibagi menjadi 2 langkah utama yang diuraikan sebagai berikut :

- 1) Menentukan faktor faktor yang mempengaruhi dengan metode Regresi Logistik Multinomial. Tahapan yang dilakukan yaitu pembentukan model awal, melakukan uji serentak dengan uji Rasio Likelihood, melakukan uji parsial dengan uji Wald dan kesimpulan.
- 2) Melakukan klasifikasi dengan menggunakan metode Regresi Logistik Multinomial dan *Fuzzy K Nearest Neighbor* (FK-NN).
 - (a) Input data.
 - (b) Membagi data menjadi data latih dan data uji. Perbandingan data latih dan data uji yang dipilih yaitu 90% : 10%.
 - (c) Klasifikasi dengan menggunakan metode Regresi Logistik Multinomial.
 - Pembentukan model klasifikasi dengan data latih.
 - Melakukan uji serentak dengan uji Rasio Likelihood.
 - Melakukan uji parsial dengan uji *Wald*.
 - Melakukan uji kesesuaian model.
 - Menentukan model akhir.
 - Pembentukan estimasi probabilitas
 - Melakukan klasifikasi menggunakan data uji untuk membentuk hasil prediksi.
 - Menghitung ketepatan klasifikasi menggunakan *Press's Q* dan APER.
 - (d) Klasifikasi dengan menggunakan *Fuzzy K Nearest Neighbor* (FK-NN).
 - Menentukan K tetangga terdekat untuk data uji x_i .
 - Menghitung nilai $u(x_i, c_j)$ atau nilai keanggotaan data uji x_i ke kelas c_j .
 - Menentukan nilai keanggotaan terbesar dari data uji x_i ke kelas c_j ($u(x_i, c_j)$).
 - Membentuk matrik konfusi dan menghitung ketepatan klasifikasinya menggunakan *Press's Q* dan APER.
 - (e) Melakukan perbandingan nilai ketepatan klasifikasi antara kedua metode (Regresi Logistik Multinomial dan *Fuzzy K Nearest Neighbor* (FK-NN)) menggunakan nilai *Press's Q* dan APER.
 - (f) Interpretasi hasil dan kesimpulan.

4. HASIL DAN PEMBAHASAN

4.1. Menentukan Faktor Faktor yang Mempengaruhi Pemilihan Metode Kontrasepsi dengan Regresi Logistik Multinomial

4.1.1. Model Tahap Pertama

Pembentukan model tahap pertama Regresi Logistik Multinomial dilakukan menggunakan variabel respon yaitu jenis metode kontrasepsi (Y) dan variabel prediktor yaitu X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, dan X13. Kategori suntik (Y(3)) pada variabel respon dijadikan sebagai kategori *baseline*. Hasil uji Rasio Likelihood yang diperoleh yaitu sebagai berikut :

Tabel 2. Output Uji Rasio Likelihood Tahap 1

llh	llhNull	G ²	$\chi^2_{(0,05:13)}$
-8682,785	-10583,45	3801,323	22,362

Berdasarkan **Tabel 2** diatas, karena nilai $G^2 = 3801,323 > \chi^2_{(0,05:13)} = 22,362$ sehingga H_0 ditolak, yang artinya paling sedikit ada satu variabel prediktor yang berpengaruh terhadap variabel respon. Sedangkan hasil uji Wald yang diperoleh yaitu sebagai berikut :

Tabel 3. Hasil Uji Wald Regresi Logistik Multinomial Tahap 1

Variabel Prediktor	Hasil Uji Wald	Kesimpulan
X1	Ada koefisien yang signifikan	Berpengaruh terhadap model
X2	Ada koefisien yang signifikan	Berpengaruh terhadap model
X3	Ada koefisien yang signifikan	Berpengaruh terhadap model
X4	Ada koefisien yang signifikan	Berpengaruh terhadap model
X5	Ada koefisien yang signifikan	Berpengaruh terhadap model
X6	Ada koefisien yang signifikan	Berpengaruh terhadap model
X7	Ada koefisien yang signifikan	Berpengaruh terhadap model
X8	Ada koefisien yang signifikan	Berpengaruh terhadap model
X9	Ada koefisien yang signifikan	Berpengaruh terhadap model
X10	Tidak ada koefisien yang signifikan	Tidak berpengaruh terhadap model
X11	Tidak ada koefisien yang signifikan	Tidak berpengaruh terhadap model
X12	Ada koefisien yang signifikan	Berpengaruh terhadap model
X13	Ada koefisien yang signifikan	Berpengaruh terhadap model

4.1.2. Model Tahap Kedua

Berdasarkan **Tabel 3**, karena variabel X10 dan X11 tidak berpengaruh terhadap model tahap pertama sehingga dilakukan pengujian kembali tanpa mengikutsertakan kedua variabel tersebut. Model baru yang terbentuk disebut model tahap 2. Pengujian parameter secara keseluruhan terhadap model tahap 2 diperoleh hasil sebagai berikut :

Tabel 4. Output Uji Rasio Likelihood Tahap 2

llh	llhNull	G ²	$\chi^2_{(0,05:11)}$
-8667,668	-10583,45	3831,558	19,675

Berdasarkan **Tabel 4** diatas, karena nilai $G^2 = 3831,558 > \chi^2_{(0,05:13)} = 19,675$ sehingga H_0 ditolak, yang artinya paling sedikit ada satu variabel prediktor yang berpengaruh terhadap variabel respon. Selanjutnya pada uji Wald terhadap model tahap 2 diperoleh hasil sebagai berikut :

Tabel 5. Hasil Uji Wald Regresi Logistik Multinomial Tahap 2

Variabel Prediktor	Hasil Uji Wald	Kesimpulan
X1	Ada koefisien yang signifikan	Berpengaruh terhadap model
X2	Ada koefisien yang signifikan	Berpengaruh terhadap model
X3	Ada koefisien yang signifikan	Berpengaruh terhadap model

X4	Ada koefisien yang signifikan	Berpengaruh terhadap model
X5	Ada koefisien yang signifikan	Berpengaruh terhadap model
X6	Ada koefisien yang signifikan	Berpengaruh terhadap model
X7	Ada koefisien yang signifikan	Berpengaruh terhadap model
X8	Ada koefisien yang signifikan	Berpengaruh terhadap model
X9	Ada koefisien yang signifikan	Berpengaruh terhadap model
X12	Ada koefisien yang signifikan	Berpengaruh terhadap model
X13	Ada koefisien yang signifikan	Berpengaruh terhadap model

Berdasarkan analisis diatas sehingga dapat disimpulkan bahwa variabel prediktor yang mempengaruhi pemilihan metode kontrasepsi masyarakat Kecamatan Bulakamba yaitu X1, X2, X3, X4, X5, X6, X7, X8, X9, X12, dan X13. Selanjutnya akan dilakukan klasifikasi menggunakan variabel prediktor tersebut.

4.2. Proses Klasifikasi

4.2.1. Pembagian Data

Pembagian data dilakukan dengan Program R dan diperoleh hasil bahwa dari total data sebesar 18543 jenis pemilihan metode kontrasepsi, akan digunakan 16713 data sebagai data latih dan 1830 sebagai data uji.

4.2.2. Klasifikasi Regresi Logistik Multinomial

1) Pembentukan Model

Pembentukan model klasifikasi dilakukan dengan variabel prediktor yaitu X1, X2, X3, X4, X5, X6, X7, X8, X9, X12, dan X13. Kategori suntik (Y(3)) pada variabel respon dijadikan sebagai kategori *baseline*.

2) Pengujian Parameter

(a) Uji Parameter secara Keseluruhan

Hasil uji Rasio Likelihood yang diperoleh terhadap model klasifikasi yang terbentuk yaitu sebagai berikut :

Tabel 6. Output Uji Rasio Likelihood Klasifikasi Regresi Logistik Multinomial

llh	llhNull	G ²	$\chi^2_{(0,05;11)}$
-7821,465	-9541,138	3439,348	19.675

Berdasarkan **Tabel 6** diatas, karena nilai $G^2 = 3439,348 > \chi^2_{(0,05;13)} = 19,675$ sehingga H_0 ditolak, yang artinya paling sedikit ada satu variabel prediktor yang berpengaruh terhadap variabel respon.

(b) Uji Parameter secara Parsial

Hasil uji Wald yang diperoleh terhadap model klasifikasi yaitu sebagai berikut :

Tabel 7. Hasil Uji Wald Klasifikasi Regresi Logistik Multinomial

Variabel Prediktor	Hasil Uji Wald	Kesimpulan
X1	Ada koefisien yang signifikan	Berpengaruh terhadap model
X2	Ada koefisien yang signifikan	Berpengaruh terhadap model
X3	Ada koefisien yang signifikan	Berpengaruh terhadap model
X4	Ada koefisien yang signifikan	Berpengaruh terhadap model
X5	Ada koefisien yang signifikan	Berpengaruh terhadap model

X6	Ada koefisien yang signifikan	Berpengaruh terhadap model
X7	Ada koefisien yang signifikan	Berpengaruh terhadap model
X8	Ada koefisien yang signifikan	Berpengaruh terhadap model
X9	Ada koefisien yang signifikan	Berpengaruh terhadap model
X12	Ada koefisien yang signifikan	Berpengaruh terhadap model
X13	Ada koefisien yang signifikan	Berpengaruh terhadap model

Berdasarkan analisis diatas dapat diketahui bahwa semua variabel prediktor yaitu X1, X2, X3, X4, X5, X6, X7, X8, X9, X12, dan X13 signifikan mempengaruhi model, sehingga model logit pada persamaan (26), (27), (28), (29), (30), (31), (32) dapat digunakan sebagai model akhir.

3) Kesesuaian Model

Hasil yang diperoleh dari uji kesesuaian model yaitu sebagai berikut :

Tabel 8. Output Uji Kesesuaian Model

<i>Hosmer and Lemeshow test</i>			
<i>Chi-square</i>	<i>Df</i>	χ^2_{tabel}	<i>P-value</i>
62,788	56	74,468	0,248

Bedasarkan **Tabel 8** dapat disimpulkan bahwa H_0 diterima karena nilai *chi-square* hitung yang diperoleh adalah 62,788 lebih kecil dari nilai χ^2_{tabel} yaitu 74,468 dan nilai p-value yang diperoleh sebesar 0,248 yang artinya lebih besar dari $\alpha = 0,05$. Hal tersebut menunjukkan bahwa dengan tingkat keyakinan 95% model yang dihasilkan sesuai.

4) Hasil Prediksi

Hasil prediksi secara kumulatif untuk setiap jenis metode kontrasepsi (kategori) dapat dilihat pada tabel berikut :

Tabel 9. Hasil Prediksi Secara Kumulatif per Kategori

	Jenis Metode Kontrasepsi (Kategori)							
	IUD (Y1)	MOP (Y2)	Suntik (Y3)	Kondom (Y4)	MOW (Y5)	Implant (Y6)	Pil (Y7)	Tradisional (Y8)
Aktual	26	5	1593	3	35	57	108	3
Prediksi	7	0	1792	0	31	0	0	0

5) Ketepatan Klasifikasi

Pada perhitungan ketepatan klasifikasi menggunakan *Apparent Error Rate* (APER) diperlukan sebuah matrik konfusi untuk mengetahui jumlah data yang terklasifikasikan benar dan jumlah data yang terklasifikasikan salah. Hasil dari matrik konfusi yang terbentuk yaitu sebanyak 1589 data terklasifikasikan benar dan 203 data terklasifikasikan salah. Nilai APER yang diperoleh dalam analisis ini yaitu sebesar 0,8825. Sedangkan nilai *Press's Q* diperoleh sebesar 9600,945. Berdasarkan hasil tersebut dapat diketahui bahwa nilai *Press's Q* $> \chi^2_{\alpha,1}$ yaitu 3,841, sehingga dapat disimpulkan bahwa dengan tingkat keyakinan 95% metode klasifikasi Regresi Logistik Multinomial akurat secara statistik.

4.2.3. Klasifikasi Fuzzy K-Nearest Neighbor (FK-NN)

Sebelum dilakukan proses prediksi dilakukan perhitungan jarak pada variabel prediktor dengan data ordinal menggunakan persamaan (18) dan pada variabel prediktor

dengan data nominal menggunakan persamaan (17) dengan menggunakan Ms. Excel. Selanjutnya proses prediksi dilakukan dengan *function* 'fknn' pada matlab yang menghasilkan nilai keanggotaan, matrik konfusi dan ketepatan klasifikasi. K yang digunakan pada penelitian ini yaitu K=1,2,..., 15. Sebagai contoh, nilai keanggotaan yang diperoleh dengan K= 1 yaitu sebagai berikut :

Tabel 10. Nilai Keanggotaan Untuk K=1

Data	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8
1	0	0	1	0	0	0	0	0
2	0	0	1	0	0	0	0	0
.
.
1830	0	0	1	0	0	0	0	0

Berdasarkan **Tabel 10**, nilai keanggotaan data uji ke-1 untuk K=1 pada kelas 1, kelas 2, kelas 4, kelas 5, kelas 6, kelas 7, dan kelas 8 adalah 0, sedangkan nilai keanggotaan pada kelas 3 yaitu 1. Sehingga hasil prediksi untuk data uji ke-1 yaitu termasuk pada kelas 3 (suntik), begitu seterusnya hingga data uji ke -1830. Hasil prediksi dengan K=1 secara kumulatif diperoleh sebagai berikut :

Tabel 11. Hasil Prediksi Kumulatif dengan K=1

	Jenis Metode Kontrasepsi (Kategori)							
	IUD (Y1)	MOP (Y2)	Suntik (Y3)	Kondom (Y4)	MOW (Y5)	Implant (Y6)	Pil (Y7)	Tradisional (Y8)
Aktual	26	5	1593	3	35	57	108	3
Prediksi	23	4	1617	1	36	47	101	1

Nilai ketepatan klasifikasi APER dan Press's Q dari K = 1,2, ..., 15 beberapa memiliki nilai yang hampir sama atau berdekatan. Nilai ketepatan klasifikasi APER dan Press's Q yang memiliki rentang cukup jauh yaitu ditampilkan sebagai berikut :

Tabel 12. Nilai APER dan Ketepatan Klasifikasi FKNN

K	APER	Ketepatan Klasifikasi	Press's Q
1	0,1885	0,8115	7884,660
4	0,1322	0,8678	9230,591
11	0,1208	0,8792	9518,014

Berdasarkan **Tabel 12** diatas dapat diketahui bahwa ketepatan klasifikasi terbaik pada klasifikasi *Fuzzy K-Nearest Neighbor* (FK-NN) yaitu dengan K=11. Selain itu nilai *Press's Q* dari semua K menunjukkan bahwa nilai *Press's Q* > $\chi_{\alpha,1}^2$ yaitu 3,841, sehingga dapat disimpulkan bahwa dengan tingkat keyakinan 95% metode klasifikasi *Fuzzy K-Nearest Neighbor* (FK-NN) akurat secara statistik.

4.3. Perbandingan Ketepatan Klasifikasi

Hasil ketepatan klasifikasi yang diperoleh dengan Regresi Logistik Multinomial dan *Fuzzy K-Nearest Neighbor* (FK-NN) yaitu sebagai berikut :

Tabel 13. Perbandingan Ketepatan Klasifikasi

Metode Klasifikasi	APER	ketepatan Klasifikasi	PressQ
Regresi Logistik Multinomial	11,75%	88,25%	9600,945
Fuzzy K-Nearest Neighbor (K=11)	12,08%	87,92%	9518,014

Berdasarkan hasil pada **Tabel 13** menunjukkan bahwa ketepatan klasifikasi dari metode Regresi Logistik Multinomial lebih baik dari *Fuzzy K-Nearest Neighbor* (FK-NN) pada pengelompokan kriteria pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan diperoleh beberapa kesimpulan sebagai berikut :

- 1) Klasifikasi Regresi Logistik Multinomial menghasilkan ketepatan klasifikasi sebesar 88,25% dan nilai *Press's Q* sebesar 9600,945 pada pengelompokan kriteria pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah.
- 2) Klasifikasi *Fuzzy K-Nearest Neighbor* (FK-NN) menghasilkan ketepatan klasifikasi sebesar 87,92% dan nilai *Press's Q* sebesar 9518,014 pada pengelompokan kriteria pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah.
- 3) Klasifikasi menggunakan Regresi Logistik Multinomial memiliki ketepatan klasifikasi yang lebih baik dari *Fuzzy K-Nearest Neighbor* (FK-NN) pada pengelompokan kriteria pemilihan metode kontrasepsi di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah.

DAFTAR PUSTAKA

- Bewick, V., Cheek, L., dan Ball, J. 2005. *Statistics Review 14: Logistic Regression*. Critical Care Vol. 9, No. 1 : Hal. 112-118.
- Hosmer, D.W., dan Lemeshow, S . 2000. *Applied Logistic Regression*. New York : John Wiley and Sons Inc.
- Johnson, R., dan Wichern, D. 2007. *Applied Multivariate Statistical Analysis*. New Jersey : Pearson Education, Inc.
- Kristiani, Y.P., Safitri, D., dan Ispiyanti, D. 2015. *Klasifikasi Kelompok Rumah Tangga di Kabupaten Blora Menggunakan Multivariate Adaptive Regression Spline (Mars) Dan Fuzzy K-Nearest Neighbor (FK-NN)*. Gaussian Vol. 4, No. 4 : Hal. 1077-1085.
- Li, D., Deogun, J.S., dan Wang, K. 2007. *Gene Function Classification Using Fuzzy K-Nearest Neighbor Approach*. Proceedings - 2007 IEEE International Conference on Granular Computing, GrC 2007 : Hal. 644-647.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : Penerbit Andi.
- Puspita, D., Suparti., dan Wilandari, Y. 2014. *Klasifikasi Tingkat Keluarga Sejahtera Dengan Menggunakan Metode Regresi Logistik Ordinal dan Fuzzy K-Nearest Neighbor*. Gaussian Vol. 3, No. 4 : Hal. 645-653.
- World Health Organization (WHO). 2016. *Selected Practice Recommendations for Contraceptive Use*. Switzerland : WHO Press.