



PERBANDINGAN *REGRESI LOGISTIK BINER* DAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)* TERHADAP FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KELAHIRAN DI KOTA KENDARI

JURNAL ILMIAH

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Statistika**

Oleh

PUTRI IKE PRATIWI

B2A219025

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2020**

PERBANDINGAN *REGRESI LOGISTIK BINER* DAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)* TERHADAP FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KELAHIRAN DI KOTA KENDARI

Putri Ike Pratiwi¹, Dr. Rochdi Wasono², Indah Manfaati Nur³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : putripratiwi1641@gmail.com

ABSTRAK

Putri Ike Pratiwi. 2020. Perbandingan *Regresi Logistik Biner* dan *Multivariate Adaptive Regression Spline (MARS)* Terhadap Faktor-Faktor yang Mempengaruhi Tingkat Kelahiran di Kota Kendari. Program Studi Statistika, Universitas Muhammadiyah Semarang. Dibimbing oleh Dr. Rochdi Wasono, M.Si. dan Indah Manfaati Nur, M.Si.

Tingkat kelahiran sangat berpengaruh dalam menangani masalah kependudukan. Jumlah kelahiran di Kota Kendari pada tahun 2017 mencapai 8.214 bayi. Angka ini menunjukkan bahwa Kota Kendari pada tahun 2017 merupakan Kota dengan jumlah kelahiran tertinggi di Sulawesi Tenggara. Adapun faktor-faktor yang mempengaruhi tingkat kelahiran di Kota Kendari yaitu umur perkawinan pertama istri, keikutsertaan program KB, tingkat pendidikan istri, status pekerjaan istri, tingkat pendidikan suami, status pekerjaan suami dan umur istri. Penelitian ini bertujuan untuk Membandingkan *Regresi Logistik Biner* dan *Multivariate Adaptive Regression Spline (MARS)* Terhadap Faktor-Faktor yang Mempengaruhi Tingkat Kelahiran di Kota Kendari. Data yang digunakan dalam penelitian ini adalah data SUSENAS tahun 2017. Hasil dari penelitian ini menunjukkan bahwa perbandingan antara kedua analisis dimana regresi logistik biner mempunyai ketepatan sebesar 77,74% sedangkan MARS mempunyai ketepatan sebesar 80,56% . Artinya pada kasus ini analisis MARS lebih baik dan lebih cocok digunakan dibandingkan regresi logistik biner.

Kata Kunci : MARS, Regresi Logistik Biner, Tingkat Kelahiran.

ABSTRACT

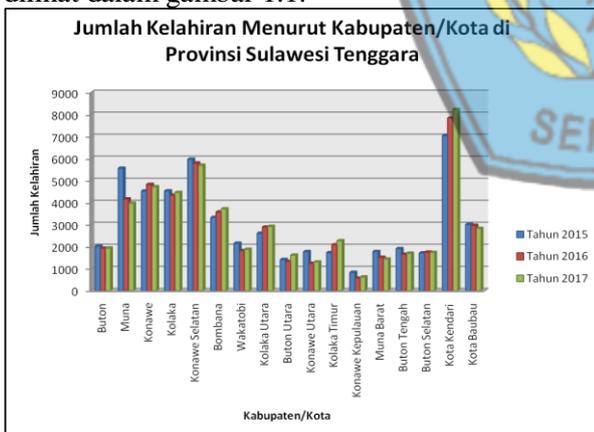
Putri Ike Pratiwi. 2020. *Comparison of Binary Logistic Regression and Multivariate Adaptive Regression Spline (MARS) Against Factors Affecting Birth Rate in Kendari City. Statistics department, Muhammadiyah University of Semarang. Supervised by Dr. Rochdi Wasono, M.Si. and Indah Manfaati Nur, M.Si.*

The birth rate is very influential in dealing with population problems. The number of births in Kendari City in 2017 reached 8,214 babies. This figure shows that Kendari City in 2017 was the city with the highest number of births in Southeast Sulawesi. The factors that affect the birth rate in Kendari are the wife's age at first marriage, participation in the family planning program, the wife's education level, the wife's employment status, the husband's education level, the husband's employment status and the wife's age. This study aims to compare the Binary Logistic Regression and the Multivariate Adaptive Regression Spline (MARS) on the factors that affect the birth rate in Kendari City. The data used in this study are SUSENAS 2017. The results of this study indicate that the comparison between the two analyzes where binary logistic regression has an accuracy of 77.74% while MARS has an accuracy of 80.56%. This means that in this case the MARS analysis is better and more suitable than binary logistic regression.

Keywords : MARS, Binary Logistic Regression, Birth Rate.

1. PENDAHULUAN

Masalah kependudukan dan lingkungan hidup merupakan tantangan Indonesia tidak terkecuali bagi Kota Kendari. Kota Kendari merupakan ibu kota Provinsi Sulawesi Tenggara. Pada tahun 2017 jumlah penduduk Kota Kendari merupakan jumlah penduduk terbanyak di Sulawesi Tenggara yaitu sebesar 370.728 jiwa, dengan jumlah penduduk terdiri dari laki-laki 187.233 jiwa dan perempuan 183.495 jiwa. Dalam rangka menangani masalah kependudukan pemerintah melakukan berbagai tindakan salah satunya adalah mengenai pengendalian kelahiran (fertilitas) dengan program keluarga berencana (KB). Fertilitas sebagai istilah demografi diartikan sebagai hasil reproduksi yang nyata dari seorang wanita (Hatmadji, 2007). Fertilitas menyangkut banyaknya bayi yang lahir hidup. Program keluarga berencana yang mulai dilaksanakan pada tahun 1970-an oleh pemerintah yang bertujuan untuk mengendalikan tingkat kelahiran di Indonesia tidak terkecuali di Kota Kendari, tingkat kelahiran sangat berpengaruh dalam menangani masalah kependudukan. Jumlah kelahiran di Kota Kendari pada tahun 2017 mencapai 8.214 bayi. Angka ini menunjukkan bahwa Kota Kendari pada tahun 2017 merupakan Kota dengan jumlah kelahiran tertinggi di Sulawesi Tenggara (BPS, 2018). Hal ini dapat dilihat dalam gambar 1.1.



Gambar 1.1 Jumlah Kelahiran Menurut Kabupaten/Kota di Provinsi Sulawesi Tenggara Tahun 2015-2017

Menurut Adioetomo (2011) banyaknya faktor yang mempengaruhi tingkat kelahiran diantaranya usia perkawinan pertama, program KB, tingkat pendidikan, dan status pekerjaan. Pembatasan jumlah anak juga perlu diperhatikan agar tercapai keluarga yang sejahtera. Salah satu

cara untuk menekan laju pertumbuhan penduduk adalah melalui program Keluarga Berencana (KB) dengan dua anak lebih baik. Oleh karena itu diperlukan suatu analisis yang berkaitan dengan faktor-faktor yang mempengaruhi tingkat kelahiran tersebut.

Banyak metode dalam ilmu statistik yang dapat digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel responnya yang bersifat kategori. Diantaranya adalah metode Regresi Logistik Biner dan *Multivariate Adaptive Regression Spline* (MARS). Kedua metode tersebut termasuk dalam analisis regresi. Menurut (Mahmuda & Karim, 2015) Regresi merupakan salah satu metode statistika yang mempelajari tentang pola hubungan secara matematis antara variabel respon (Y) dengan salah satu atau lebih variabel prediktor (X) jika variabel terikat bersifat kategori, maka dapat dianalisis menggunakan metode regresi logistik. Metode regresi logistik merupakan suatu metode yang digunakan untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih variabel bebas (Hosmer & Lemeshow, 2000). Pada penelitian ini variabel terikat terdiri dari dua kategori, sehingga metode regresi logistik yang tepat untuk digunakan adalah regresi logistik biner. Regresi logistik biner merupakan metode yang digunakan untuk menganalisis hubungan antara variabel terikat yang bersifat kategori biner dengan sejumlah variabel bebas.

Metode statistika lainnya yang dapat digunakan juga untuk pengklasifikasian objek dan melihat hubungan antara variabel respon yang bersifat kategori dengan sejumlah variabel bebas adalah *Multivariate Adaptive Regression Spline* (MARS). MARS merupakan metode yang tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar dalam penggunaannya. Metode ini bertujuan untuk mengatasi permasalahan data berdimensi tinggi dan menghasilkan variabel respon yang akurat, serta memperbaiki kelemahan dengan menghasilkan model yang kontinu dalam *knot* berdasarkan nilai *generalized cross validation* terkecil (Friedman, 1991).

Berdasarkan permasalahan yang dikemukakan di atas, maka dilakukan penelitian yang berjudul “Perbandingan *Regresi Logistik Biner* dan *Multivariate Adaptive Regression Spline* (MARS) terhadap Faktor-faktor yang Mempengaruhi Tingkat Kelahiran di Kota Kendari”.

2. TINJAUAN PUSTAKA

2.1 Regresi Logistik Biner

Regresi logistik merupakan analisis yang penting dalam melihat hubungan antara suatu variabel respon dengan satu atau lebih variabel penjelas dengan data kategorik.

Apabila variabel respon terdiri atas dua kategori yaitu $Y = 1$ (Sukses) dan $Y = 0$ (Gagal), metode regresi logistik yang dapat diterapkan adalah regresi logistik biner. Variabel yang dikotomik/biner adalah variabel yang hanya mempunyai dua kategori saja, yaitu kategori yang menyatakan kejadian sukses ($Y = 1$) dan kategori yang menyatakan kejadian gagal ($Y = 0$). Variabel terikat Y ini, diasumsikan mengikuti distribusi Bernoulli.

2.2 Bentuk Umum Model Regresi Logistik

Bentuk umum model peluang regresi logistik adalah sebagai berikut:

$$P(Y=1) = \pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}$$

dimana:

X_i : Variabel penjelas, dengan $i = 1, 2, \dots, p$.

$\pi(x)$: Peluang terjadinya kejadian yang sukses yaitu $Y = 1$.

β_j : Nilai parameter ke- j ($j = 0, 1, 2, \dots, p$).

Persamaan tersebut merupakan model nonlinear maka perlu ditransformasikan ke dalam bentuk *logit* agar dapat dilihat hubungan variabel penjelas dengan variabel respon (Hosmer & Lemeshow, 2000).

2.3 Bentuk Linear dari Regresi Logistik

Dengan melakukan transformasi logit dari $\pi(x)$ diperoleh persamaan yang lebih sederhana yang merupakan fungsi linear yaitu:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right]$$

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Persamaan tersebut merupakan fungsi linear dalam parameter-parameternya atau $g(x)$ telah linear dalam parameter-parameternya (Cullagh, 1989).

2.4 Penaksiran Parameter

Model regresi logistik menggunakan metode *Maximum Likelihood* untuk menduga parameter-parameternya. Dalam model regresi logistik variabel terikat mengikuti distribusi

Bernoulli dengan fungsi kepadatan peluang sebagai berikut:

$$f(\beta, y_i) = \pi(x)^{y_i} [1 - \pi(x)]^{1-y_i}$$

Karena nilai variabel terikat (Y_i) diasumsikan saling bebas, maka diperoleh fungsi *likelihood* sebagai berikut:

$$l(\beta) = \prod_{i=1}^k f(\beta, y_i)$$

Sedangkan persamaan *likelihood* adalah:

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

Untuk mendapatkan nilai β yang memaksimumkan $L(\beta)$ dilakukan diferensiasi

(turunan) terhadap β , dengan syarat: (1) $\frac{\partial L}{\partial \beta} = 0$,

dan (2) $\frac{\partial^2 L}{\partial \beta^2} < 0$

Sehingga diperoleh persamaan:

$$\sum_{i=1}^k x [y_i - \pi(x)] = 0$$

Persamaan tidak linear dalam β , sehingga solusi bagi $\hat{\beta} = \hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ tidak dapat dituliskan secara eksplisit. Untuk mendapatkan nilai β digunakan metode iterasi *Newton Raphson*. Iterasi merupakan metode yang umum dalam membantu perhitungan estimasi dari β (Hosmer & Lemeshow, 2000).

2.5 Uji Signifikansi Parameter

2.5.1 Uji Simultan

Uji yang digunakan untuk menguji kelayakan model dengan menggunakan keseluruhan variabel penjelas secara bersama-sama digunakan uji *likelihood ratio* atau uji signifikansi model.

Hipotesis:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0$ (tidak ada pengaruh antara variabel-variabel penjelas terhadap variabel respon).

H_1 : Minimal ada satu $\beta_j \neq 0$ (minimal ada satu variabel penjelas yang berpengaruh terhadap variabel respon).

Statistik uji:

$$G = -2 \ln \left[\frac{(L_0)}{(L_1)} \right]$$

dimana:

L_0 : fungsi *likelihood* tanpa variabel penjelas.

L_1 : fungsi *likelihood* dengan semua variabel penjelas.

Statistik G mengikuti sebaran khi-kuadrat dengan derajat bebas banyaknya parameter yang terdapat dalam model ($k-1$) (Agresti, 1990).

Kriteria uji:

$G > \chi^2_{(df;\alpha)}$ atau $p\text{-value} < \alpha$, maka H_0 ditolak.

$G \leq \chi^2_{(df;\alpha)}$ atau $p\text{-value} \geq \alpha$, maka H_0 diterima.

2.5.2 Uji Parsial

Uji yang digunakan untuk mengetahui variabel penjelas yang mempengaruhi variabel respon, digunakan uji koefisien parameter β secara parsial dengan menggunakan uji wald.

Hipotesis:

$H_0 : \beta_j = 0$ (tidak ada pengaruh antara variabel penjelas ke- j dengan variabel respon)

$H_1 : \beta_j \neq 0$ (ada pengaruh antara variabel penjelas ke- j dengan variabel respon)

Statistik uji:

$$W = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right)^2$$

dimana:

$\hat{\beta}_j$: penduga β_j

$se(\hat{\beta}_j)$: galat baku dari $\hat{\beta}_j$

Kriteria uji:

Kriteria uji yang digunakan berdasarkan statistik uji W diasumsikan mengikuti sebaran khi-kuadrat dengan derajat bebas = 1 (Agresti, 1990).

Jika $W > \chi^2_{(df;\alpha)}$ atau $p\text{-value} < \alpha$, maka H_0 ditolak.

Jika $W \leq \chi^2_{(df;\alpha)}$ atau $p\text{-value} \geq \alpha$, maka H_0 diterima.

2.6 Uji Kesesuaian Model

Untuk menguji kesesuaian model, dilakukan uji Hosmer Lemeshow terhadap model regresi logistik.

Hipotesis:

H_0 : Model sesuai/tepat (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

H_1 : Model tidak sesuai/tepat (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

Kriteria uji:

Jika $p\text{-value} < \alpha$, maka H_0 ditolak.

Jika $p\text{-value} \geq \alpha$, maka H_0 diterima.

2.7 Odds Ratio

Odds ratio merupakan perbandingan tingkat resiko relatif dari dua buah nilai variabel penjelas X_j atau resiko kecenderungan misalkan $X_j = 1$ terhadap $X_j = 0$. Dengan kata lain, resiko kecenderungan pengaruh observasi $X_j = 1$ adalah n kali lipat dibandingkan observasi $X_j = 0$. Odds ratio dilambangkan dengan ψ yang merupakan ukuran untuk mengetahui tingkat resiko, yaitu perbandingan antara dua nilai variabel penjelas X_j antara kejadian-kejadian yang masuk kategori sukses dan gagal.

Persamaan dari odds ratio (Hosmer & Lemeshow, 2000) dituliskan sebagai berikut:

$$\psi = \frac{\pi(1)/1-\pi(1)}{\pi(0)/1-\pi(0)}$$

dimana: (2.7)

Untuk $X_j = 1$, digunakan:

$$\frac{\pi(1)}{1-\pi(1)} = \frac{\exp\{\beta_0 + \beta_j\}}{1 + \exp\{\beta_0 + \beta_j\}} = \frac{\exp\{\beta_0 + \beta_j\}}{1} = \exp\{\beta_0 + \beta_j\}$$

Untuk $X_j = 0$, digunakan:

$$\frac{\pi(0)}{1-\pi(0)} = \frac{\exp\{\beta_0\}}{1 + \exp\{\beta_0\}} = \frac{\exp\{\beta_0\}}{1 + \exp\{\beta_0\} - \exp\{\beta_0\}} = \exp\{\beta_0\}$$

Sehingga diperoleh persamaan sebagai berikut:

$$\begin{aligned} \psi &= \frac{\pi(1)/1-\pi(1)}{\pi(0)/1-\pi(0)} \\ &= \frac{\exp\{\beta_0 + \beta_j\}}{\exp\{\beta_0\}} \\ &= \exp\{\beta_0 + \beta_j - \beta_0\} \\ &= \exp\{\beta_j\} \end{aligned}$$

2.8 Multivariate Adaptive Regression Spline (MARS)

MARS merupakan salah satu pendekatan regresi nonparametrik multivariat yang berguna untuk mengatasi permasalahan data yang

berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebanyak $3 \leq n \leq 20$ (Friedman, 1991). Selain itu model MARS mampu menghasilkan prediksi variabel respon yang akurat. Metode MARS menjadi populer karena tidak menentukan tipe khusus seperti hubungan (linear, kuadratik, dan kubik) diantara variabel prediktor dan respon pada proses pembentukan model MARS tidak memerlukan asumsi (Otok, Guritno, Subanar & Haryatmi, 2006).

Beberapa istilah yang perlu diperhatikan dalam pemodelan MARS adalah sebagai berikut.

1. Knot

Knot merupakan suatu nilai/titik tempat perubahan pola apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data yang ada dari variabel prediktor. Knot merupakan akhir dari sebuah garis regresi dan juga awal dari garis regresi yang lain (Nash & Bradford, 2001).

2. Basis Fungsi

Basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Basis fungsi bisa memiliki lebih dari satu variabel yang merupakan fungsi dari tiap garis regresi yang dihasilkan. Maksimum basis fungsi yang diperbolehkan adalah 2 sampai 4 kali jumlah variabel prediktornya.

3. Interaksi

Interaksi merupakan hubungan korelasi antar variabel dengan maksimum interaksi (MI) adalah 1, 2, dan 3. Jika MI lebih dari tiga maka akan menghasilkan model yang lebih kompleks.

4. Minimum Observasi

Minimum Observasi (MO) merupakan jumlah pengamatan paling minimal antar knot sebesar 0, 1, 2, dan 3.

Secara umum estimator model MARS dapat ditulis pada persamaan berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m \prod_{k=1}^{k_m} [S_{km}(X_{v(k,m)} - t_{km})]$$

dengan :

$\hat{\alpha}_0$ = koefisien konstanta basis fungsi B_0

α_m = koefisien dari basis fungsi ke- m

M = maksimum basis fungsi

m = banyaknya basis fungsi

k_m = banyaknya interaksi pada basis fungsi m

k = banyaknya interaksi

S_{km} = nilainya 1 atau -1 jika data berada di sebelah kanan atau kiri titik knot

$X_{v(k,m)}$ = variabel prediktor

t_{km} = nilai knot dari variabel prediktor $x_{v(k,m)}$

Berdasarkan persamaan tersebut model MARS dapat dituliskan sebagai berikut.

$$y_i = \alpha_0 + \sum_{m=1}^M \alpha_m B_{im}(x) + \varepsilon_i$$

dengan

$B_{im}(x) = \prod_{k=1}^{k_m} [S_{km}(X_{v(k,m)} - t_{km})]$.
Sehingga jika ditulis dalam bentuk matriks dapat menjadi

$$y = B \alpha + \varepsilon$$

dengan,

$$y = (y_1 y_2 y_3 \dots y_n)^T, \alpha = (\alpha_1 \alpha_2 \alpha_3 \dots \alpha_M)^T, \varepsilon = (\varepsilon_1 \varepsilon_2 \varepsilon_3 \dots \varepsilon_n)^T$$

$$B = \begin{pmatrix} 1 & \prod_{k=1}^{k_1} [S_{k1}(x_{v(k,1)} - t_{k1})] & \dots & \prod_{k=1}^{k_M} [S_{kM}(x_{v(k,M)} - t_{kM})] \\ 1 & \prod_{k=1}^{k_1} [S_{k1}(x_{v(k,1)2} - t_{k1})] & \dots & \prod_{k=1}^{k_M} [S_{kM}(x_{v(k,M)2} - t_{kM})] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{k_1} [S_{k1}(x_{v(k,1)n} - t_{k1})] & \dots & \prod_{k=1}^{k_M} [S_{kM}(x_{v(k,M)n} - t_{kM})] \end{pmatrix}$$

Persamaan tersebut dapat dijabarkan sebagai berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(X_{v(1m)} - t_{1m})] + \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(X_{1m} - t_{1m})] + [S_{2m}(X_{v(2m)} - t_{2m})] + \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(X_{v(1m)} - t_{1m})] + [S_{2m}(X_{v(2m)} - t_{2m})] + [S_{3m}(X_{v(3m)} - t_{3m})] + \dots$$

Secara umum persamaan tersebut dapat dituliskan sebagai berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{i=1}^p \hat{f}_i(x_i) + \sum_{i,j=1}^p \hat{f}_{ij}(x_i, x_j) + \sum_{i,j,k=1}^p \hat{f}_{ijk}(x_i, x_j, x_k) + \dots$$

Persamaan tersebut menunjukkan bahwa penjumlahan suku pertama meliputi semua basis fungsi untuk satu variabel. Penjumlahan suku kedua meliputi semua basis fungsi untuk interaksi antara dua variabel. Penjumlahan suku ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel, dan seterusnya.

Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai GCV yang minimum (Nisa & Budiantara, 2010). Pemilihan model pada MARS dapat menggunakan metode *stepwise* (*forward* dan *backward*). Pemilihan model dengan menggunakan *forward stepwise* dilakukan untuk mendapatkan jumlah basis fungsi maksimum, sedangkan pada *backward stepwise* dilakukan pemilihan basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross Validation* (GCV). Model terbaik dipilih berdasarkan nilai GCV yang paling

minimum. Fungsi GCV minimum didefinisikan sebagai berikut.

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_m(x_i)]^2}{\left[1 - \frac{c(M)}{N}\right]^2}$$

dengan,

N : banyaknya pengamatan

y_i : variabel respon

x_i : variabel prediktor

$\hat{f}_m(x_i)$: nilai taksiran variabel respon pada pengamatan ke- i

$c(M)$: jumlah parameter dalam model = $Trace [B(B^T B)^{-1} B^T] + 1$

B : matriks basis fungsi

2.9 Klasifikasi MARS Respon Biner

Klasifikasi pada model MARS didasarkan pada analisis regresi. Jika variabel respon terdiri dari dua nilai, maka dikatakan sebagai regresi dengan respon biner (Cox & Snell, 1989), sehingga dapat digunakan model probabilitas dengan persamaan sebagai berikut.

$$P(Z = 1 | X = x) = \pi(x) = \frac{e^{f(x)}}{1 + e^{f(x)}}$$

dan

$$P(Z = 0 | X = x) = 1 - \pi(x) = \frac{1}{1 + e^{f(x)}}$$

Dengan $f(x) = z = \text{logit } \pi(x)$. Model MARS untuk klasifikasi dapat dinyatakan sebagai berikut.

$$z = \text{logit } \pi(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [S_{km}(x_{v(km)} - t_{km})]$$

2.10 Ketepatan Klasifikasi

Ketepatan klasifikasi diperlukan untuk mengetahui pengelompokan data yang digolongkan dengan tepat pada kelompoknya. *Apparent Error Rate* (APER) didefinisikan sebagai proporsi sampel yang tidak tepat diklasifikasikan (Johnson & Wichern, 2007). Untuk mengetahui proporsi sampel yang tepat diklasifikasikan dapat dihitung dari nilai TAR (*Total Accuracy Rate*). Berikut ini merupakan tabel pengklasifikasian untuk respon biner.

Tabel 2.1 Klasifikasi Respon Biner

| Observasi | Taksiran Observasi | |
|-----------|--------------------|----------|
| | y_0 | y_1 |
| y_0 | n_{00} | n_{01} |
| y_1 | n_{10} | n_{11} |

Nilai APER dan TAR didapatkan dengan perhitungan sebagai berikut.

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \times 100\%$$

dan

$$TAR(\%) = 1 - APER = 1 - \left(\frac{n_{10} + n_{01}}{n} \times 100\%\right)$$

dengan,

N = jumlah observasi

n_{00} = jumlah observasi dari y_0 yang tepat diklasifikasikan sebagai y_0

n_{11} = jumlah observasi dari y_1 yang tepat diklasifikasikan sebagai y_1

n_{01} = jumlah observasi dari y_0 yang salah diklasifikasikan sebagai y_1

n_{10} = jumlah observasi dari y_1 yang salah diklasifikasikan sebagai y_0

3. METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) 2017 untuk Kota Kendari. Jumlah penduduk Kota Kendari tahun 2017 yaitu sebanyak 370.728 jiwa, dengan jumlah penduduk terdiri dari laki-laki sebanyak 187.233 jiwa dan perempuan sebanyak 183.495 jiwa dengan Wanita Pernah Kawin (WPK) yang menjadi sampel SUSENAS 2017 di Kota Kendari yang berumur 15-49 tahun sebanyak 319 sampel.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari dua jenis variabel antara lain variabel terikat (*dependen*) dan variabel bebas (*independen*). Variabel-variabel tersebut disajikan pada tabel berikut:

Tabel 3.1 Variabel Penelitian

| Nama Variabel | Keterangan | Skala Data |
|---|---|------------|
| Jumlah anak yang dilahirkan (Y) | 1 = jumlah anak yang dilahirkan ≤ 2 0 = jumlah anak yang dilahirkan > 2 | Nominal |
| Umur perkawinan pertama istri (X ₁) | 1 = umur perkawinan pertama ≥ 21 tahun 0 = umur perkawinan pertama < 21 tahun | Nominal |
| Keikutsertaan program KB (X ₂) | 1 = menggunakan KB 0 = tidak menggunakan KB | Nominal |
| Tingkat pendidikan istri (X ₃) | 1 = SD 2 = SMP dan SMA 3 = Perguruan Tinggi | Ordinal |
| Status pekerjaan istri (X ₄) | 1 = bekerja 0 = tidak bekerja | Nominal |
| Tingkat pendidikan suami (X ₅) | 1 = SD 2 = SMP dan SMA 3 = Perguruan Tinggi | Ordinal |
| Status pekerjaan suami (X ₆) | 1 = bekerja 0 = tidak bekerja | Nominal |
| Umur istri (X ₇) | Umur istri | Rasio |

3.3 Analisis Data

Untuk mencapai tujuan penelitian, maka langkah-langkah yang dilakukan dalam penelitian yaitu:

1. Mendeskripsikan data responden
2. Melakukan analisis regresi logistik dengan langkah-langkah sebagai berikut:
 - a. Menghitung estimasi parameter
 - b. Melakukan uji signifikan parameter secara simultan
 - c. Melakukan uji signifikansi parameter secara parsial
 - d. Menentukan model akhir dari variabel yang berpengaruh
 - e. Melakukan uji kesesuaian model
 - f. Menginterpretasikan model
 - g. Menguji ketepatan klasifikasi

3. Melakukan analisis *MARS* dengan langkah-langkah sebagai berikut:
 - a. Pembentukan model *MARS* untuk data set awal
 - b. Menentukan jumlah basis fungsi, maksimum interaksi, dan minimum observasi (MO)
 - c. Mendapatkan model *MARS* terbaik untuk data set berdasarkan nilai GCV terkecil
 - d. Mengklasifikasi kedalam bentuk matriks konfusi
 - e. Menginterpretasikan model
 - f. Menguji ketepatan klasifikasi
4. Membandingkan ketepatan klasifikasi model analisis *Regresi Logistik Biner* dan *Multivariate Adaptive Regression Spline (MARS)*
5. Menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif

Berikut adalah analisis karakteristik responden mengenai faktor-faktor yang mempengaruhi tingkat kelahiran terhadap jumlah anak yang dimiliki.

4.1.1 Umur Perkawinan Pertama Istri

Tabel 4.1 Umur perkawinan pertama istri

| Umur Perkawinan Pertama Istri (X ₁) | Jumlah Anak | | Total | Persentase | |
|---|-------------|-------|-------|------------|-------|
| | ≤ 2 | > 2 | | ≤ 2 | > 2 |
| < 21 | 58 | 74 | 132 | 43.94 | 56.06 |
| ≥ 21 | 114 | 73 | 187 | 60.96 | 39.04 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.1 bahwa wanita yang umurnya pada saat perkawinan pertamanya kurang dari 21 tahun cenderung memiliki jumlah anak lebih dari 2 sedangkan untuk wanita yang umur perkawinan pertamanya lebih atau sama dengan 21 tahun cenderung memiliki jumlah anak kurang dari atau sama dengan 2.

4.1.2 Keikutsertaan Program Keluarga Berencana (KB)

Tabel 4.2 Keikutsertaan Program KB

| Keikutsertaan Program KB (X ₂) | Jumlah Anak | | Total | Persentase | |
|--|-------------|-------|-------|------------|-------|
| | ≤ 2 | > 2 | | ≤ 2 | > 2 |
| Tidak Menggunakan KB | 99 | 72 | 171 | 57.89 | 42.11 |
| Menggunakan KB | 73 | 75 | 148 | 49.32 | 50.68 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.2 bahwa wanita yang tidak menggunakan KB cenderung memiliki jumlah anak kurang dari atau sama dengan 2 sedangkan untuk wanita yang menggunakan KB cenderung memiliki jumlah anak lebih dari 2.

4.1.3 Tingkat Pendidikan Istri

Tabel 4.3 Tingkat pendidikan istri

| Tingkat Pendidikan Istri (X3) | Jumlah Anak | | Total | Persentase | |
|-------------------------------|-------------|-----|-------|------------|-------|
| | ≤2 | >2 | | ≤2 | >2 |
| SD | 21 | 24 | 45 | 46.67 | 53.33 |
| SMP & SMA | 102 | 88 | 190 | 53.68 | 46.32 |
| Perguruan Tinggi | 49 | 35 | 84 | 58.33 | 41.67 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.3 bahwa wanita dengan tingkat pendidikan SD cenderung memiliki jumlah anak lebih dari 2, untuk wanita dengan tingkat pendidikan SMP & SMA cenderung memiliki jumlah anak kurang dari atau sama dengan 2 sedangkan untuk wanita dengan tingkat pendidikan Perguruan Tinggi cenderung memiliki jumlah anak kurang dari atau sama dengan 2.

4.1.4 Status Pekerjaan Istri

Tabel 4.4 Status pekerjaan istri

| Status Pekerjaan Istri (X4) | Jumlah Anak | | Total | Persentase | |
|-----------------------------|-------------|-----|-------|------------|-------|
| | ≤2 | >2 | | ≤2 | >2 |
| Tidak Bekerja | 83 | 61 | 144 | 57.64 | 42.36 |
| Bekerja | 89 | 86 | 175 | 50.86 | 49.14 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.4 bahwa wanita yang tidak bekerja dan bekerja cenderung memiliki jumlah anak kurang dari atau sama dengan 2.

4.1.5 Tingkat Pendidikan suami

Tabel 4.5 Tingkat pendidikan suami

| Tingkat Pendidikan suami (X5) | Jumlah Anak | | Total | Persentase | |
|-------------------------------|-------------|-----|-------|------------|-------|
| | ≤2 | >2 | | ≤2 | >2 |
| SD | 15 | 22 | 37 | 40.54 | 59.46 |
| SMP & SMA | 116 | 78 | 194 | 59.79 | 40.21 |
| Perguruan Tinggi | 41 | 47 | 88 | 46.59 | 53.41 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.5 bahwa pria dengan tingkat pendidikan SD cenderung memiliki jumlah anak lebih dari 2, untuk pria dengan tingkat pendidikan SMP & SMA cenderung memiliki jumlah anak kurang dari atau sama dengan 2 sedangkan untuk pria dengan tingkat pendidikan Perguruan Tinggi cenderung memiliki jumlah anak lebih dari 2.

4.1.6 Status Pekerjaan Suami

Tabel 4.6 Status pekerjaan suami

| Status Pekerjaan Ayah (X6) | Jumlah Anak | | Total | Persentase | |
|----------------------------|-------------|-----|-------|------------|-------|
| | ≤2 | >2 | | ≤2 | >2 |
| Tidak Bekerja | 6 | 7 | 13 | 46.15 | 53.85 |
| Bekerja | 166 | 140 | 306 | 54.25 | 45.75 |
| Total | 172 | 147 | 319 | 53.92 | 46.08 |

Terlihat dari tabel 4.6 bahwa pria yang tidak bekerja cenderung memiliki jumlah anak lebih dari 2 sedangkan untuk pria yang bekerja cenderung memiliki jumlah anak kurang dari atau sama dengan 2.

4.1.7 Umur Istri

Tabel 4.7 Umur Istri

| Variabel | Jumlah anak yang dimiliki | Mean | Standar Deviasi | Max | Min |
|--------------|---------------------------|-------|-----------------|-----|-----|
| Umur (Tahun) | Jumlah Anak ≤ 2 | 32,45 | 7,25 | 49 | 17 |
| | Jumlah Anak > 2 | 40,78 | 5,91 | 49 | 24 |

Berdasarkan Tabel 4.7 diketahui bahwa rata-rata umur istri yang memiliki jumlah anak ≤ 2 yaitu ± 32 tahun sedangkan rata-rata umur istri yang memiliki jumlah anak > 2 yaitu ± 40 tahun. Adapun standar deviasi umur istri yang memiliki jumlah anak ≤ 2 sebesar 7,25 sedangkan standar deviasi umur istri yang memiliki jumlah anak > 2 sebesar 5,91. Selain itu, diketahui juga bahwa umur istri paling muda yang memiliki jumlah anak ≤ 2 yaitu 17 tahun sedangkan umur paling tua yaitu 49 tahun. Kemudian umur istri paling muda yang memiliki jumlah anak > 2 yaitu 24 tahun sedangkan umur paling tua yaitu 49 tahun.

4.2 Analisis Regresi Logistik Biner

Analisis data yang pertama dilakukan adalah menggunakan regresi logistik biner. Langkah awal dalam metode ini adalah melakukan estimasi parameter.

4.2.1. Estimasi Parameter

Untuk mendapatkan model awal, diperlukan perhitungan estimasi parameter. Nilai estimasi parameter yang dihasilkan adalah sebagai berikut:

Tabel 4.8 Estimasi Parameter Model Awal

| Variabel Prediktor (X_j) | $\hat{\beta}_j$ |
|---|-----------------|
| Konstanta(β_0) | 6.75471 |
| Umur Perkawinan Pertama Istri (X_1) | 1.42762 |
| Keikutsertaan Program KB (X_2) | -0.89975 |
| Tingkat Pendidikan Istri (X_3_2) | 0.01332 |
| Tingkat Pendidikan Istri (X_3_3) | 0.17160 |
| Status Pekerjaan Istri (X_4) | 0.07599 |
| Tingkat Pendidikan Suami (X_5_2) | 0.86584 |
| Tingkat Pendidikan Suami (X_5_3) | 0.49614 |
| Status Pekerjaan Suami (X_6) | 0.10137 |
| Umur Istri (X_7) | -0.21357 |

Dengan melihat tabel di atas, maka model awal yang akan terbentuk adalah :

$$g(x) = 6.75471 + 1.42762X_1 - 0.89975X_2 + 0.01332X_{32} + 0.17160X_{33} + 0.07599X_4 + 0.86584X_{52} + 0.49614X_{53} + 0.10137X_6 - 0.21357X_7$$

4.2.2. Uji Signifikansi Parameter

Setelah mendapatkan estimasi parameter, selanjutnya adalah melakukan uji signifikansi parameter baik secara bersama-sama ataupun masing-masing prediktor.

1. Uji Rasio Likelihood

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_7 = 0$ (secara bersama-sama variabel prediktor tidak mempengaruhi model)

$H_1 : \text{minimal ada satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, 7. \text{ (secara bersama-sama variabel prediktor mempengaruhi model).}$

Taraf signifikansi : $\alpha = 5\%$

Statistik Uji

$$G = 2 \ln \left[\frac{\text{likelihood tanpa variabel penjelas}}{\text{likelihood dengan variabel penjelas}} \right] = 140,100$$

Kriteria uji : H_0 ditolak jika $G > X^2_{(0,05;7)}$

Keputusan : Karena $(G = 140,100) > (X^2_{(0,05;7)} = 14,067)$ maka H_0 ditolak

Kesimpulan :

Jadi, pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa variabel prediktor secara bersama-sama mempengaruhi model.

2. Uji Wald

Hipotesis :

$H_0 : \beta_j = 0$ (variabel ke- j tidak mempengaruhi model)

$H_1 : \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, 7. \text{ (variabel ke- j mempengaruhi model)}$

Taraf signifikansi : $\alpha = 5\%$

Statistik Uji

$$w = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right)^2$$

Kriteria uji : H_0 ditolak jika $W > X^2_{(0,05;1)}$

Nilai Wald untuk masing-masing variabel adalah sebagai berikut:

Tabel 4.9 Nilai Wald Model Awal

| Variabel Prediktor | Wald | $X^2_{(0,05;1)}$ | Keputusan |
|---|--------|------------------|----------------|
| Konstanta | 28.058 | 3,841 | |
| Umur Perkawinan Pertama Istri (X_1) | 18.905 | 3,841 | H_0 ditolak |
| Keikutsertaan Program KB (X_2) | 8.679 | 3,841 | H_0 ditolak |
| Tingkat Pendidikan Istri (X_{32}) | 0.001 | 3,841 | H_0 diterima |
| Tingkat Pendidikan Istri (X_{33}) | 0.092 | 3,841 | H_0 diterima |
| Status Pekerjaan Istri (X_4) | 0.064 | 3,841 | H_0 diterima |
| Tingkat Pendidikan Suami (X_{52}) | 3.059 | 3,841 | H_0 diterima |
| Tingkat Pendidikan Suami (X_{53}) | 0.692 | 3,841 | H_0 diterima |
| Status Pekerjaan Suami (X_6) | 0.018 | 3,841 | H_0 diterima |
| Umur Istri (X_7) | 73.599 | 3,841 | H_0 ditolak |

Kesimpulan :

Jadi, pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa variabel yang berpengaruh terhadap model adalah variabel umur perkawinan pertama istri (X_1), keikutsertaan program KB (X_2) dan umur istri (X_7), sedangkan variabel lainnya tidak berpengaruh terhadap model.

Selanjutnya perlu dilakukan pembentukan model baru dengan menggunakan variabel yang hanya berpengaruh terhadap model yaitu variabel umur perkawinan pertama istri (X_1), keikutsertaan program KB (X_2) dan umur istri (X_7).

Berikut ini merupakan uji rasio likelihood dan uji wald terhadap variabel prediktor tersebut.

1. Uji Rasio Likelihood

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \beta_7 = 0$ (secara bersama-sama variabel prediktor tidak mempengaruhi model)

$H_1 : \beta_1 = \beta_2 = \beta_7 \neq 0$ (secara bersama-sama variabel prediktor mempengaruhi model).

Taraf signifikansi : $\alpha = 5\%$

Statistik Uji

$$G = 2 \ln \left[\frac{\text{likelihood tanpa variabel penjelas}}{\text{likelihood dengan variabel penjelas}} \right] = 136,053$$

Kriteria uji : H_0 ditolak jika $G > X^2_{(0,05;3)}$

Keputusan : Karena $(G = 136,053) > (X^2_{(0,05;3)} = 7,815)$ maka H_0 ditolak

Kesimpulan :

Jadi, pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa variabel prediktor mempengaruhi model.

2. Uji Wald

Hipotesis :

$H_0 : \beta_j = 0$ (variabel ke- j tidak mempengaruhi model)

$H_1 : \beta_j \neq 0$, untuk $j = 1, 2, 7$. (variabel ke- j mempengaruhi model)

Taraf signifikansi : $\alpha = 5\%$

Statistik Uji

$$W = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right)^2$$

Kriteria uji : H_0 ditolak jika $W > X^2_{(0,05;1)}$

Nilai Wald untuk masing-masing variabel adalah sebagai berikut

Tabel 4.10 Nilai Wald Untuk Model Akhir

| Variabel Prediktor | $\hat{\beta}_j$ | Wald | $X^2_{(0,05;1)}$ | Keputusan |
|------------------------------------|-----------------|--------|------------------|---------------|
| Konstanta | 7.54767 | 67.289 | 3,841 | |
| Umur Perkawinan Pertama (X_1) | 1.49111 | 22.724 | 3,841 | H_0 ditolak |
| Keikutsertaan Program KB (X_2) | -0.84367 | 8.083 | 3,841 | H_0 ditolak |
| Umur Istri (X_7) | -0.21369 | 77.475 | 3,841 | H_0 ditolak |

Kesimpulan :

Jadi, pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa variabel Umur Perkawinan Pertama Istri (X_1), Keikutsertaan Program KB (X_2) dan Umur Istri (X_7) berpengaruh terhadap model.

4.2.3. Uji Godness of Fit

Hipotesis :

H_0 : tidak ada perbedaan antara prediksi dan hasil observasi (model sesuai)

H_1 : ada perbedaan antara prediksi dan hasil observasi (model tidak sesuai)

Taraf signifikansi : $\alpha = 5\%$

Statistik Uji

$$\hat{C} = \sum_{k=1}^g \frac{(\sigma_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} = 10.036$$

Kriteria uji : H_0 ditolak jika $\hat{C} > X^2_{(0,05;7)}$

Keputusan : Karena $(\hat{C} = 10.036) < (X^2_{(0,05;7)} = 14,067)$ maka H_0 diterima

Kesimpulan :

Jadi, pada taraf signifikansi $\alpha = 5\%$ dapat disimpulkan bahwa tidak ada perbedaan antara prediksi dan hasil observasi. Ini berarti model yang diperoleh sudah sesuai.

4.2.4. Model Akhir

Model akhir yang terbentuk adalah sebagai berikut.

$$g(x) = 7.54767 + 1.49111X_1 - 0.84367X_2 - 0.21369 X_7$$

Interpretasi model akhir yang terbentuk dilihat dari nilai odds rasionya. Untuk X_1 (Umur Perkawinan Pertama Istri), nilai odds rasionya adalah $e^{1.49111} = 4,44$. Artinya perempuan yang dengan umur perkawinan pertama < 21 tahun lebih beresiko 4,44 kali memiliki jumlah anak yang dilahirkan > 2 dibandingkan dengan jumlah anak yang dilahirkan ≤ 2 .

Untuk X_2 (Keikutsertaan Program KB), nilai odds rasionya adalah $e^{-0.843} = 0,43$. Artinya perempuan yang menggunakan KB lebih beresiko 0,43 kali memiliki jumlah anak yang dilahirkan > 2 dibandingkan dengan jumlah anak yang dilahirkan ≤ 2 .

Untuk X_7 (Umur Istri), nilai odds rasionya adalah $e^{-0.21369} = 0,81$. Artinya setiap penambahan satu tahun umur perempuan maka akan menurunkan resiko sebesar 0,81 kali memiliki jumlah anak yang dilahirkan > 2 .

4.2.5. Klasifikasi

Berikut ini merupakan hasil ketepatan klasifikasi dari metode regresi logistik biner.

Tabel 4.11 Ketepatan Klasifikasi Logistik Biner

| Hasil Observasi | Hasil Prediksi | |
|----------------------|-------------------|----------------------|
| | Jumlah Anak > 2 | Jumlah Anak ≤ 2 |
| Jumlah Anak > 2 | 109 | 38 |
| Jumlah Anak ≤ 2 | 33 | 139 |

$$APER (\%) = \frac{n_{10} + n_{01}}{n} \times 100\% = \frac{38 + 33}{319} \times 100\% = 22,26\%$$

$$TAR (\%) = 1 - APER = 1 - 22,26\% = 77,74\%$$

Nilai TAR sebesar 77,74% menunjukkan bahwa model logistik biner yang diperoleh telah mampu memprediksi dengan tepat sebesar 77,74 %.

4.3 Analisis MARS

Pembentukan model MARS dapat dilakukan dengan cara *trial and error* untuk semua kombinasi BF, MI, dan MO. Jumlah BF yang digunakan adalah 2 sampai 4 kali jumlah variabel prediktor. Pada penelitian ini jumlah variabel prediktor yang digunakan sebanyak 7 sehingga jumlah BF yang digunakan adalah 14, 21, dan 28. MI yang digunakan sebanyak 1, 2, dan 3. Sedangkan MO yang digunakan sebanyak 0, 1, 2, dan 3. Berikut ini merupakan hasil *trial and error* untuk setiap kombinasi BF, MI, dan MO

Tabel 4.12 Trial and Error Pembentukan Model Mars

| No | BF | MI | MO | GCV |
|----|----|----|----|---------|
| 1 | 14 | 1 | 0 | 0.15271 |
| 2 | 14 | 1 | 1 | 0.15223 |
| 3 | 14 | 1 | 2 | 0.15223 |
| 4 | 14 | 1 | 3 | 0.15223 |
| 5 | 14 | 2 | 0 | 0.15099 |
| 6 | 14 | 2 | 1 | 0.15099 |
| 7 | 14 | 2 | 2 | 0.15099 |
| 8 | 14 | 2 | 3 | 0.15099 |
| 9 | 14 | 3 | 0 | 0.15019 |
| 10 | 14 | 3 | 1 | 0.15019 |
| 11 | 14 | 3 | 2 | 0.15019 |
| 12 | 14 | 3 | 3 | 0.15019 |
| 13 | 21 | 1 | 0 | 0.14644 |
| 14 | 21 | 1 | 1 | 0.14609 |
| 15 | 21 | 1 | 2 | 0.14609 |
| 16 | 21 | 1 | 3 | 0.14609 |
| 17 | 21 | 2 | 0 | 0.14449 |
| 18 | 21 | 2 | 1 | 0.14656 |
| 19 | 21 | 2 | 2 | 0.14669 |
| 20 | 21 | 2 | 3 | 0.14701 |
| 21 | 21 | 3 | 0 | 0.14644 |
| 22 | 21 | 3 | 1 | 0.14474 |
| 23 | 21 | 3 | 2 | 0.14474 |
| 24 | 21 | 3 | 3 | 0.14553 |
| 25 | 28 | 1 | 0 | 0.14447 |
| 26 | 28 | 1 | 1 | 0.14411 |
| 27 | 28 | 1 | 2 | 0.14411 |
| 28 | 28 | 1 | 3 | 0.14411 |
| 29 | 28 | 2 | 0 | 0.14154 |
| 30 | 28 | 2 | 1 | 0.14399 |
| 31 | 28 | 2 | 2 | 0.14401 |
| 32 | 28 | 2 | 3 | 0.14479 |
| 33 | 28 | 3 | 0 | 0.14035 |
| 34 | 28 | 3 | 1 | 0.14148 |
| 35 | 28 | 3 | 2 | 0.14148 |
| 36 | 28 | 3 | 3 | 0.14271 |

Berdasarkan Tabel 4.12 diketahui bahwa jumlah model yang terbentuk untuk setiap kombinasi adalah 36 model. Pemilihan model terbaik dapat dilihat dari nilai GCV yang terkecil. Dari 36 model tersebut, model yang memiliki nilai GCV terkecil yaitu model ke-33 dengan jumlah BF sebanyak 28, jumlah MI sebanyak 3, dan jumlah MO sebanyak 0.

4.3.1. Model Terbaik

Model terbaik adalah model yang dihasilkan dari nilai GCV terkecil. Jadi model yang dihasilkan dari kombinasi BF, MI, dan MO tersebut merupakan model terbaik. Model dari kombinasi BF = 28, MI = 3 dan MO = 0 adalah sebagai berikut:

$$\hat{f}(x) = 0,236 + 0,402BF1 - 0,130BF2 - 0,188BF3 + 0,674BF4 - 0,617BF5 + 0,028BF6 + 0,034BF7 + 0,188BF8 - 0,162BF9 + 0,620BF10 - 0,134BF11 - 0,014BF12 - 0,060BF13 + 0,053BF14 - 0,486BF15 - 0,694BF16$$

Dengan

$$BF1 = X_1$$

$$BF2 = X_2$$

$$BF3 = \max(0, X_7 - 38)$$

$$BF4 = \max(0, X_7 - 40)$$

$$BF5 = \max(0, X_7 - 41)$$

$$BF6 = \max(0, X_7 - 42)$$

$$BF7 = \max(0,44 - X_7)$$

$$BF8 = \max(0, X_7 - 44)$$

$$BF9 = \max(0, X_7 - 46)$$

$$BF10 = X_1 * \max(0, X_5 - 2)$$

$$BF11 = X_1 * \max(0, 2 - X_5)$$

$$BF12 = X_1 * \max(0, X_7 - 30)$$

$$BF13 = X_1 * \max(0, 30 - X_7)$$

$$BF14 = X_1 * \max(0, X_7 - 44)$$

$$BF15 = X_1 * X_2 * \max(0, 2 - X_5)$$

$$BF16 = X_1 * \max(0, X_5 - 2) * X_6$$

Dari model terbaik yang dihasilkan, dapat disimpulkan bahwa variabel Umur Perkawinan Pertama Istri (X_1), Keikutsertaan Program KB (X_2), Tingkat Pendidikan Suami (X_5), Status Pekerjaan Suami (X_6), Umur Istri (X_7) berpengaruh terhadap tingkat kelahiran.

4.3.2. Klasifikasi

Berikut ini merupakan hasil ketepatan klasifikasi dari metode MARS.

Tabel 4.13 Ketepatan Klasifikasi MARS

| Hasil Observasi | Hasil Prediksi | |
|-----------------|----------------|-----------------|
| | Jumlah Anak >2 | Jumlah Anak ≤ 2 |
| Jumlah Anak >2 | 120 | 27 |
| Jumlah Anak ≤ 2 | 35 | 137 |

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \times 100\% = \frac{27 + 35}{319} \times 100\% = 19,44$$

$$TAR(\%) = 1 - APER = 1 - 19,44\% = 80,56\%$$

Nilai TAR sebesar 80,56% menunjukkan bahwa model MARS yang diperoleh telah mampu memprediksi dengan tepat sebesar 80,56%.

4.4 Perbandingan Metode Logistik Biner dan MARS

Tabel 4.14 Perbandingan Ketepatan Klasifikasi

| Ketepatan | Persentase | |
|-----------|----------------|--------|
| | Logistik Biner | MARS |
| APER | 22,26% | 19,44% |
| TAR | 77,74% | 80,56% |

Berdasarkan tabel di atas, dapat diketahui bahwa ketepatan klasifikasi menggunakan metode MARS lebih tinggi dibandingkan Regresi Logistik Biner. Dimana persentase ketepatan klasifikasi MARS dilihat dari persentase (TAR) adalah sebesar 80,56%, sedangkan pada ketepatan klasifikasi Regresi Logistik Biner adalah sebesar 77,74%.

5. KESIMPULAN

1. WPK yang memiliki yang memiliki jumlah anak yang dilahirkan lebih dari dua sebanyak 147 orang dan yang memiliki jumlah anak kurang dari atau sama dengan dua sebanyak 172 orang.

2. a. Model yang diperoleh menggunakan analisis regresi logistik biner adalah sebagai berikut.

$$\hat{g}(x) = 7.54767 + 1.49111X_1 - 0.84367X_2 - 0.21369 X_7$$

dengan faktor-faktor yang mempengaruhi tingkat kelahiran di Kota Kendari yaitu variabel Umur Perkawinan Pertama Istri (X_1), Keikutsertaan Program KB (X_2) dan Umur Istri (X_7).

- b. Model yang diperoleh menggunakan analisis MARS adalah sebagai berikut.

$$\hat{f}(x) = 0,236 + 0,402BF1 - 0,130BF2 - 0,188BF3 + 0,674BF4 - 0,617BF5 + 0,028BF6 + 0,034BF7 + 0,188BF8 - 0,162BF9 + 0,620BF10 - 0,134BF11 - 0,014BF12 - 0,060BF13 + 0,053BF14 - 0,486BF15 - 0,694BF16$$

dengan faktor-faktor yang mempengaruhi tingkat kelahiran di Kota Kendari yaitu variabel Umur Perkawinan Pertama Istri (X_1), Keikutsertaan Program KB (X_2), Tingkat Pendidikan Suami (X_5), Status Pekerjaan Suami (X_6) dan Umur Istri (X_7).

3. Ketepatan klasifikasi menggunakan metode MARS lebih tinggi dibandingkan Regresi Logistik Biner. Dimana persentase ketepatan klasifikasi MARS dilihat dari persentase (TAR) adalah sebesar 80,56%, sedangkan pada ketepatan klasifikasi Regresi Logistik Biner adalah sebesar 77,74%. Hal ini menunjukkan bahwa metode MARS lebih cocok digunakan pada kasus faktor-faktor yang mempengaruhi tingkat kelahiran di Kota kendari dibandingkan dengan metode Regresi Logistik Biner.

DAFTAR PUSTAKA

- Adioetomo, Moertiningsih, S, Samosir, & Bulan, O. 2011. *Dasar-Dasar Demografi*. Jakarta Selatan: Penerbit Salemba Empat.
- Agresti, A. 1990. *Categorical Data Analysis*. New York: John & Willey Sons.
- BPS Provinsi Sulawesi Tenggara. 2018. *Sulawesi Tenggara Dalam Angka 2018*. Sulawesi Tenggara.
- Cox, D. R., & Snell, E. J. (1989). *Analysis of Binary Data* (2nd ed.). London: Chapman & Hall/CRC.
- Cullagh, 1989. *Generalized Linear Models 2nd Edition*. London: Chapman Hall.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1-67.

Hatmadji, Sri. 2007. *Dasar-dasar Demografi*. Jakarta: Lembaga Demografi FEUI.

Hosmer, D.W & Lemeshow, S. 2000. *Applied Logistic Regression*. New York: John Willeyand Sons, Inc.

Johnson, R. A. & Wichern, D. W. 1992. *Applied Logistic Regression*. USA : John Wiley & Sons.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. United States of America: Pearson Prentice Hall.

Lembaga Demografi UI. 2010. *Dasar-Dasar Demografi*. Salemba Empat. Jakarta.

Lucas, David. 1990. *Pengantar Kependudukan*. Cetakan Keempat. Yogyakarta : Gajah Mada Universitas Press.

Manginsihi, (2003). *Pengertian pekerjaan*. <https://digilib.unimed.ac.id>[8 maret 2016].

Mantra, IB. 2003. *Demografi Umum*. Yogyakarta: Pustaka Pelajar.

Nash, M. S., & Bradford, D. F. (2001). *Parametric and Non Parametric Logistic Regression for Prediction of Precense/Absence of an Amphibian*. Las Vegas: Nevada.

Nisa', S. S., & Budiantara, I. N. (2012). Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD). *Jurnal Sains dan Seni ITS*.

Otok, B. W., Guritno, S., Subanar, & Haryatmi, S. (2006). Bootsrap dalam MARS untuk Klasifikasi Perbankan. *Inferensi Jurnal Statistik*.

Rajagukguk, N., Ispriyanti, D. & Wilandari, Y. 2015. *Perbandingan Metode Klasifikasi Regresi Logistik Biner Dan Naive Bayes Pada Status Pengguna KB Di Kota Tegal Tahun 2014*. *Jurnal Gaussian* 4(2): 365 – 374.

Yahya, Irma dkk. 2019. Pemodelan Keberhasilan Program Keluarga Berencana (KB) di Kota Kendari dengan pendekatan Regresi Logistik Biner. Seminar Nasional Teknologi Terapan Inovasi dan Rekayasa (SNT2IR) Program Pendidikan Vokasi Universitas Halu Oleo, 210-215.