



**PERBANDINGAN METODE KLASIFIKASI PADA FAKTOR- FAKTOR
YANG MEMPENGARUHI PARTISIPASI TENAGA KERJA DI
KABUPATEN TANAH DATAR, SUMATERA BARAT**

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

**NADIA KEMALA GUSTI
B2A219034**

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

PERSETUJUAN PEMBIMBING

Skripsi dengan judul **“Perbandingan Metode Klasifikasi pada Faktor-Faktor yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat”** yang disusun oleh:

Nama : Nadia Kemala Gusti
NIM : B2A219034
Program Studi : S1-Statistika

Telah disetujui oleh dosen pembimbing pada tanggal : 17 September 2020



PENGESAHAN KELULUSAN

Skripsi dengan judul “Perbandingan Metode Klasifikasi pada Faktor-Faktor yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat” yang disusun oleh:

Nama : Nadia Kemala Gusti
NIM : B2A219034
Program Studi : S1-Statistika

Telah dipertahankan dalam Sidang Panitia Ujian Skripsi Program Sarjana Universitas Muhammadiyah Semarang Pada Tanggal : 17 September 2020

Panitia Ujian

Ketua Tim Penguji


Dr. Rochdi Wasono, M.Si

NIK. 28.6.1026.119

Anggota Tim Penguji I


Prizka Rismawati Arum, M.Stat.

CP. 1026.071

Anggota Tim Penguji II


Tiani Wahyu Utami, M.Si

NIK. 28.6.1026.341

Anggota Tim Penguji III


Fatkhurrohman Fauzi, M.Stat

Mengetahui,

Ketua Program Studi


Indah Manraati Nur, M.Si

NIK. 28.6.1026.221

**SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH**

Nama : Nadia Kemala Gusti
NIM : B2A219034
Fakultas/ Jurusan : S1-Statistika
Jenis Penelitian : Skripsi
Judul : Perbandingan Metode Klasifikasi pada Faktor - Faktor yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat
Email : nadiakemala23@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih mediakan/mengalih formatkan, mengelola dalam bentuk pengalalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada perpustakaan Unimus, tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta.
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 17 September 2020
Yang menyatakan



Nadia Kemala Gusti
NIM. B2A219034

PERBANDINGAN METODE KLASIFIKASI PADA FAKTOR- FAKTOR YANG MEMPENGARUHI PARTISIPASI TENAGA KERJA DI KABUPATEN TANAH DATAR, SUMATERA BARAT

Nadia Kemala Gusti¹, Tiani Wahyu Utami², Fatkhurrohman Fauzi³

^{1,2,3}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : nadiakemala23@gmail.com

ABSTRAK

Tenaga kerja adalah salah satu penentu kesuksesan sebuah perekonomian yang diisi oleh Partisipasi Tenaga Kerja yang tepat. Dimana dengan partisipasi tenaga kerja yang tepat dapat mengoptimalkan hasil luaran dari perekonomian tersebut. Salah satunya adalah Tanah datar, Sumatera barat. Provinsi Sumatera Barat merupakan sebuah provinsi di Indonesia yang berkembang di bidang tenaga kerja. Tujuan dari penelitian ini adalah untuk Membandingkan *Regresi Logistik Biner*, *Multivariate Adaptive Regression Spline (Mars)* dan *Bootstrap Aggregating Multivariate Adaptive Regression Splines (Bagging MARS)* terhadap Faktor-faktor Yang Mempengaruhi Partisipasi Tenaga Kerja di Tanah Datar, Sumatera Barat. Data yang digunakan dalam penelitian ini adalah data sekunder tahun 2019 yang diperoleh dari Survei Angkatan Kerja Nasional (SAKERNAS), Badan Pusat Statistik. Metode pertama yang digunakan dalam analisis ini adalah Regresi Logistik Biner. Hasil analisis dari model terbaik diketahui bahwa variabel Pendidikan (X_5) dan status dalam RT (X_6) adalah variabel yang signifikan. Metode kedua yang digunakan adalah MARS. Pada metode ini model terbaik diperoleh dari nilai GCV terkecil, dimana diketahui variabel yang signifikan adalah Usia (X_1), Kesehatan (X_2), Jaminan Sosial (X_3), Jenis Kelamin (X_4), dan Status dalam RT (X_6). Metode ketiga yang digunakan adalah Bagging MARS yaitu metode untuk meningkatkan akurasi model MARS. Berdasarkan hasil analisis didapatkan hasil perbandingan antara ketiga metode dimana regresi logistik biner mempunyai ketepatan sebesar 68,45%, MARS sebesar 82,74% dan Bagging MARS sebesar 83,33%. Artinya pada kasus ini analisis Bagging MARS lebih baik dan lebih cocok digunakan dibandingkan regresi logistik biner dan MARS.

Kata Kunci : Regresi Logistik Biner, MARS, Bagging MARS, Partisipasi Tenaga Kerja.

ABSTRACT

Labor is one of the determinants of the success of an economy filled with appropriate labor participation. Where with the right labor participation can optimize the output of the economy. One of them is flat land, West Sumatra. West Sumatra Province is a province in Indonesia that develops in the field of labor . The purpose of this study was to compare *Binary Logistic Regression* , *Multivariate Adaptive Regression Spline (Mars)* and *Bootstrap aggregating Multivariate Adaptive Regression Splines (MARS Bagging)* against Factors That Affect Labor Force Participation in Tanah Datar, West Sumatra. The data used in the study of this is the data secondary years 2019 were obtained from the Survey Force Employment National (SAKERNAS), Agency Center for Statistics . The first method used in this analysis is Binary Logistic Regression. The results of the analysis of the best model show that the education variable (X_5) and the status in RT (X_6) are significant variables. The second method used is MARS. In this method the best model is obtained from the smallest GCV value, where it is known that the significant variables are Age (X_1) , Health (X_2) , Social Security (X_3), Gender (X_4) and Status in RT (X_6). The third method used is Bagging MARS, which is a method to increase the accuracy of the MARS model. Based on the results of the analysis of obtained results of the comparison between all three methods in which regression logistic binary has a precision of 68,45% , MARS amounted to 82.74% and amounted to 83.33% MARS Bagging. This means that in the case of this analysis Bagging MARS is better and more suitable use than regression logistic binary and MARS .

Keywords : *Binary Logistic Regression, Multivariate Adaptive Regression Spline (MARS), MARS Bagging, Workforce Participation*

PENDAHULUAN

Tenaga kerja adalah salah satu penentu kesuksesan sebuah perekonomian. Tenaga kerja dalam masyarakat merupakan faktor yang potensial untuk pembangunan ekonomi secara keseluruhan (Swasono, 1983). Sektor perekonomian yang diisi oleh partisipasi tenaga kerja yang tepat. Dimana dengan partisipasi tenaga kerja yang tepat dapat mengoptimalkan hasil luaran dari perekonomian tersebut. Begitu pentingnya peran partisipasi tenaga kerja dalam menjalankan perekonomian.

Salah satunya adalah Tanah Datar, Sumatera Barat. Sumatera Barat merupakan sebuah provinsi di Indonesia yang berkembang di bidang tenaga kerja. Namun menurut Badan Pusat Statistik Provinsi Sumatera Barat periode Februari 2019 jumlah angkatan kerja Provinsi Sumatera Barat sebanyak 2,69 juta jiwa mengalami penurunan dibandingkan tahun 2018 sebanyak 2,74 juta jiwa. Sedangkan pada Februari 2019 tingkat partisipasi angkatan kerja Provinsi Sumatera Barat 70,27% mengalami penurunan dibandingkan tahun sebelumnya yaitu 72,80%.

Menurut Badan Pusat Statistik (BPS), salah satu indikator tenaga kerja adalah berkeja dan tidak bekerja. Menurut Raviantor (1985:14) ada beberapa faktor yang mempengaruhi seseorang untuk bekerja diantaranya 1). Pendidikan, 2). Kesehatan, 3). Jaminan Kesehatan. Adapun faktor lain yang mempengaruhi partisipasi tenaga kerja adalah jenis kelamin, usia dan status dalam rumah tangga.

Analisis statistik yang dapat digunakan untuk mengetahui pengaruh variabel bebas terhadap variabel terikat adalah analisis regresi. Regresi logistik biner merupakan metode yang digunakan untuk menganalisis hubungan antara variabel terikat yang bersifat kategori biner dengan sejumlah variabel bebas. Kajian mengenai klasifikasi menggunakan metode Regresi Logistik Biner pernah dilakukan oleh Hasnita (2020) untuk memprediksi interaksi farmakodinamik. Hasil yang didapatkan bahwa metode Regresi Logistik Biner memberikan hasil akurasi yang lebih baik dibandingkan metode lainnya sebesar 63,15%.

Metode statistika lainnya yang dapat digunakan untuk melihat hubungan antara variabel respon yang bersifat kategori dengan sejumlah variabel bebas adalah *Multivariate Adaptive Regression Spline* (MARS). Tingkat akurasi klasifikasi model MARS dapat ditingkatkan menggunakan *resampling*, salah

satunya dengan metode *Bootstrap Aggregating* (Bagging) yang dapat memperbaiki stabilitas, meningkatkan akurasi, dan kekuatan prediktif (Breiman, 1994).

Kajian mengenai klasifikasi menggunakan metode MARS pernah dilakukan oleh Binadari (2015) untuk membandingkan peminatan jurusan SMA Negeri 2 Semarang. Hasil yang didapatkan bahwa metode MARS memberikan hasil akurasi yang lebih baik dibandingkan metode lainnya sebesar 83,38%. Penelitian selanjutnya menurut Rahmaniah (2016) menggunakan metode *Bootstrap Aggregating Multivariate Adaptive Regression Splines* (Bagging MARS) untuk mengidentifikasi komponen pencari akreditasi Sekolah/Madrasah pada tingkat SD/MI di Provinsi Kalimantan Timur. Hasil yang didapatkan bahwa metode Bagging MARS memberikan hasil akurasi sebesar 89,44%.

Berdasarkan permasalahan di atas pada penelitian ini akan dibahas tentang “Perbandingan Regresi Logistik Biner, *Multivariate Adaptive Regression Splines* (MARS) dan *Bootstrap Aggregating Multivariate Adaptive Regression Splines* (Bagging MARS) terhadap Faktor-faktor Yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat”, dimana faktor-faktornya adalah pendidikan, kesehatan, jaminan sosial, jenis kelamin, usia dan status kepala rumah tangga.

TINJAUAN PUSTAKA

1. Analisis Regresi Logistik Biner Model Regresi Logistik Biner

Analisis regresi logistik merupakan salah satu metode regresi yang dapat digunakan untuk menggambarkan hubungan variabel terikat (Y) yang bersifat kategorik dengan satu atau lebih variabel bebas (X) yang bersifat kontinu, kategori atau kombinasi keduanya (Agresti, 2002:165). Model regresi logistik biner dengan k variabel prediktor adalah sebagai berikut:

$$P(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + (\exp \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}$$

Persamaan di atas mempunyai bentuk yang tidak linier. Untuk membuat persamaan tersebut menjadi persamaan yang linier, maka digunakan transformasi log atau disebut juga transformasi logit. Berikut ini adalah logit dari $P(x)$

$$g(x) = \text{logit}(P(x)), \text{ dimana } P(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

$$g(x) = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

Pengujian Signifikansi Parameter

Berikut ini adalah pengujian signifikansi parameter yaitu :

a. Uji Rasio Likelihood

Hipotesis :

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

H_1 : paling sedikit ada satu $\beta_i \neq 0$, untuk $i = 1, 2, 3, \dots, k$

Dengan statistika uji

$$G = 2 \ln \left[\frac{\text{Likelihood tanpa variabel penjelas}}{\text{Likelihood tanpa variabel penjelas}} \right]$$

Kriteria uji : Tolak H_0 jika $G > \chi_{\alpha, k}^2$ atau nilai signifikansi kurang dari α .

b. Uji Wald

Hipotesis:

$H_0: \beta_j = 0$, untuk $j = 1, 2, \dots, k$ (peubah X_j tidak berpengaruh nyata)

$H_1: \beta_j \neq 0$ (peubah X_j berpengaruh nyata)

Dengan statistik uji :

$$W_j = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}; j = 0, 1, 2, \dots, k$$

Tolak H_0 jika $W > \chi_{\alpha, 1}^2$ atau nilai signifikansi kurang dari α .

c. Uji Kesesuaian Model

Pengujian kesesuaian model dilakukan menggunakan *Hosmer-Lemeshow goodness-of-fit test* dengan hipotesis sebagai berikut :

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

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

Statistik uji :

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

Tolak H_0 jika $\chi_{hitung}^2 > \chi_{(db, \alpha)}^2$ atau nilai signifikansi kurang dari α .

2. Multivariate Adaptive Regression Spline

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). Secara umum estimator model MARS dapat ditulis pada persamaan berikut.

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} [S_{km}(X_{v(km)} - t_{km})]$$

dengan :

α_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

$X_{v(km)}$ = variabel prediktor

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

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

a. 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.

b. 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.

c. Minimum Observasi

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

3. Bagging MARS

Bootstrap Aggregating (Bagging) merupakan teknik yang diusulkan oleh Breiman (1996) yang dapat digunakan untuk me-reduksi variansi estimator pada metode klasifikasi dan regresi. Penggunaannya tidak dibatasi hanya untuk memperbaiki estimator. Teknik ini juga dapat memperbaiki stabilitas, meningkatkan akurasi, dan kekuatan prediktif. Bagging merupakan salah satu bagian dari *Bootstrap*. Perbedaan pada Bagging dilakukan dengan membandingkan sampel secara berpasangan.

Buhlmann dan Yu (2002), secara singkat menyatakan algoritma Bagging sebagai berikut.

1. Sebuah data set L yang terdiri dari $\{(y_i, x_i), i = 1, 2, \dots, n\}$. Melakukan replikasi bootstrap pada data, sehingga didapatkan $L_i^* = (y_i^*, x_i^*)$, dengan $i = 1, 2, \dots, n$.
2. Replikasi *bootstrap* dilakukan sebanyak B kali, sehingga didapatkan $(L^{(B)})$ yang merupakan *resampling* pengembalian dari L .

4. Ketetapan 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 dan 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

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

METODE PENELITIAN

1. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder tahun 2019, yang diperoleh dari Survei Angkatan Kerja Nasional (SAKERNAS), Badan Pusat Statistik sebanyak 168 responden.

2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari dua jenis variabel antara lain variabel prediktor dan variabel respon. Variabel-variabel tersebut disajikan pada tabel berikut:

Tabel 3.1 Variabel Penelitian

| Nama Variabel | Keterangan |
|------------------------------|----------------------------------|
| Partisipasi Tenaga Kerja (Y) | 1 = tidak bekerja 0 = bekerja |
| Usia (X_1) | |
| Kesehatan (X_2) | 1 = tidak sehat 0 = sehat |

| | |
|--------------------------|--|
| Jaminan Sosial (X_3) | 1 = tidak ada 0 = ada |
| Jenis Kelamin (X_4) | 1 = laki-laki 0 = perempuan |
| Pendidikan (X_5) | 2 = pendidikan tinggi 1 = pendidikan menengah 0 = pendidikan dasar |
| Staus dalam RT (X_6) | 1 = bukan kepala RT 0 = kepala RT |

3. Analisis Data

Berikut merupakan langkah-langkah analisis penelitian yang akan dilakukan.

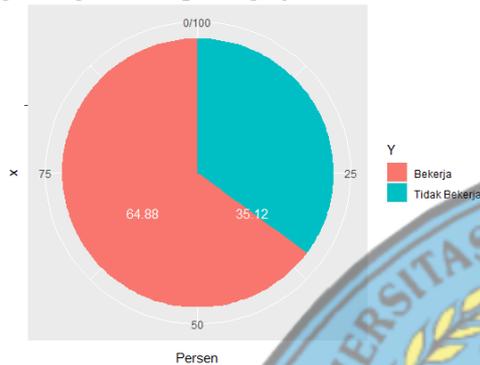
1. Mengumpulkan data partisipasi tenaga kerja.
2. Analisis statistika deskriptif untuk mengetahui karakteristik dari partisipasi tenaga kerja.
3. Melakukan analisis regresi logistik biner dengan menggunakan software R
4. Menghitung estimasi parameter
5. Melakukan uji signifikansi dengan uji likelihood dan uji wald
6. Menentukan model akhir dari variabel yang paling berpengaruh
7. Melakukan uji kesesuaian model
8. Menguji ketepatan klasifikasi
9. Melakukan analisis *MARS* menggunakan software R
10. Pembentukan model *MARS* untuk data set awal
11. Menentukan jumlah basis fungsi (BF), maksimum interkasi (MI), dan minimum observasi (MO). Jumlah BF yang digunakan adalah 2 sampai 4 kali jumlah variabel prediktor yaitu 12, 18, dan 24. Banyaknya MI pada model yang digunakan adalah 1, 2, dan 3. MO yang digunakan pada penelitian ini adalah 0, 1, 2 dan 3.
12. Mendapatkan model *MARS* terbaik untuk data set berdasarkan nilai GCV terkecil
13. Menginterpretasikan model
14. Menguji ketepatan klasifikasi
15. Melakukan analisis *Bagging MARS* menggunakan software R
16. Melakukan pemodelan *Bagging MARS* pada data dengan replikasi 50, 100, 150, 200 dan 500 kali.
17. Memilih akurasi ketepatan klasifikasi terbaik dari hasil replikasi metode *Bagging MARS* berdasarkan nilai presentase keakuratan tertinggi

18. Untuk menjawab tujuan dilakukan dengan membandingkan tingkat akurasi klasifikasi antara metode Logistik Biner, MARS dan Bagging MARS untuk mendapatkan metode terbaik.
19. Meginterpretasi hasil analisis.

HASIL DAN PEMBAHASAN

1. Eksporasi Data

Karakteristik partisipasi tenaga kerja dapat dilihat dari analisis deskriptif pada masing-masing variabel. Berikut statistika deskriptif untuk variabel partisipasi tenaga kerja yaitu:



Gambar 4.1 Persentase Partisipasi Tenag Kerja

Dapat diketahui pada Gambar 4.1 yang menunjukkan persentase partisipasi tenaga kerja berdasarkan bekerja atau tidak bekerjanya masyarakat di Kabupaten Tanah Datar, Sumatera Barat. Hasil klasifikasi menunjukkan bahwa 64,88% atau sejumlah 109 orang yang bekerja dan 35,12% atau sejumlah 59 orang yang tidak bekerja. Dalam hal ini, dapat dikatakan bahwa orang yang bekerja lebih banyak dibanding orang yang tidak bekerja. Adanya perbedaan hasil tersebut diharapkan dapat diketahui faktor-faktor apa yang menyebabkan partisipasi tenaga kerja dan hasil akurasi klasifikasi yang tepat berdasarkan faktor-faktornya.

2. Analisis Regresi Logistik Biner

Analisis regresi logistik dilakukan pada data dependen bersifat kategori, yang tujuannya untuk menentukan variabel Usia (X_1), Kesehatan (X_2), Jaminan Sosial (X_3), Jenis Kelamin (X_4), Pendidikan (X_5) dan Status dalam RT (X_6). yang signifikan mempengaruhi partisipasi tenaga kerja. Berikut adalah pendugaan parameter regresi logistik biner.

Estimasi Parameter

Model regresi logistik dibentuk setelah melakukan pendugaan parameter regresi logistik

terlebih dahulu yang dilakukan menggunakan MLE (*Maximum Likelihood Estimator*) dengan mengikutsertakan semua variabel bebas. Berikut ini merupakan hasil pendugaan parameter model regresi logistik biner

$$g(x) = -17,207 - 0,002X_1 + 0,376X_2 + 15,622X_3 - 0,130X_4 - 0,443X_5 + 1,872X_6$$

Model yang diperoleh perlu dipertimbangkan kembali sehingga perlu dilakukan pengujian signifikansi model regresi logistik.

Uji Signifikansi Parameter

Setelah mendapatkan estimasi parameter, selanjutnya adalah melakukan uji signifikansi parameter baik secara bersama-sama menggunakan uji rasio likelihood ataupun masing-masing prediktor menggunakan uji wald. Berikut adalah hasil uji rasio likelihood dan uji wald yaitu :

a. Uji Rasio Likelihood dan Uji Wald

Berikut hasil uji signifikansi model dengan menggunakan uji G, Untuk melihat pengaruh variabel bebas yang terlibat pada model yang setelah proses reduksi. Berdasarkan data partisipasi tenaga kerja dilihat pada Tabel 4.1 berikut

Tabel 4.1 Nilai G Untuk Model Akhir

| | χ^2 | $\chi^2_{(0,05;1)}$ |
|------------------------|----------|---------------------|
| Model Regresi Logistik | 25,144 | 3,841 |

Berdasarkan Tabel 4.1 di atas, nilai statistik uji G hasil reduksi adalah 25,144 yang berarti bahwa model reduksi yang diperoleh sama baiknya dengan model yang melibatkan seluruh variabel bebas. Selanjutnya untuk hasil uji wald adalah sebagai berikut:

Tabel 4.2 Nilai Wald Untuk Model Akhir

| Variabel Prediktor | Wald | p-value |
|---------------------------|--------|-------------------------|
| Pendidikan (X_5) | 5,371 | 0,0217 |
| Status dalam RT (X_6) | 17,501 | 4,6548X10 ⁻⁵ |

Berdasarkan Tabel 4.2 di atas, terlihat bahwa variabel Pendidikan(X_5) dan Status dalam RT (X_6) memberikan pengaruh yang signifikan terhadap variabel terikat dimana memiliki nilai signifikansi lebih kecil dari taraf nyata 0,05.

b. Uji Godness of Fit

Berikut hasil uji kesesuaian model regresi logistik biner dengan menggunakan uji *Godness of Fit* berdasarkan data partisipasi tenaga kerja dilihat pada Tabel 4.3 berikut

Tabel 4.3 Nilai Kecocokan Untuk Model Akhir

| χ^2 | $\chi^2_{(0,05;8)}$ |
|----------|---------------------|
| 0,20545 | 15,507 |

Berdasarkan tabel 4.3 di atas diketahui bahwa nilai $(X^2 = 0,20545) < (X^2_{(0,05;8)} = 15,507)$ maka H_0 diterima sehingga dapat disimpulkan bahwa tidak ada perbedaan antara prediksi dan hasil observasi. Ini berarti model yang diperoleh sudah sesuai.

Model Akhir

Setelah melakukan uji kecocokan model, maka didapatkan model regresi logistik biner yang menggambarkan faktor yang mempengaruhi partisipasi tenaga kerja di Kabupaten Tanah Datar, Sumatera Barat. Berikut hasil dari model akhir regresi logistik biner

Tabel 4.4 Nilai Estimasi Parameter Model Akhir

| Variabel Prediktor | $\hat{\beta}_j$ |
|---------------------------|-----------------|
| Konstanta | -1,6227 |
| Pendidikan (X_5) | -0,6285 |
| Status dalam RT (X_6) | 1,8244 |

Berdasarkan hasil pada Tabel 4.4 di atas, diperoleh model regresi logistik biner bebas sebagai berikut:

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

Dengan:

$$g(x) = -1,6227 - 0,6285X_5 + 1,8244X_6$$

Interpretasi model akhir yang terbentuk dilihat dari nilai odds rasionya. Untuk Pendidikan (X_5) nilai odds rasionya adalah $e^{-0,6285} = 0,5$. Artinya setiap kenaikan satu satuan Pendidikan (X_5), maka partisipasi tenaga kerja seseorang yang bekerja adalah 0,5 kali dari pada tidak bekerja. Sedangkan untuk Status dalam RT (X_6), nilai odds rasionya adalah $e^{1,8244} = 6,2$. Artinya setiap kenaikan satu satuan Status dalam RT (X_6), maka partisipasi tenaga kerja seseorang yang bekerja adalah 6,2 kali dari pada tidak bekerja.

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 6 sehingga jumlah BF yang digunakan adalah 12, 18, dan 24. MI yang digunakan sebanyak 1, 2, dan 3. Sedangkan MO yang digunakan sebanyak 0, 1, 2, dan 3.

Berdasarkan 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, R^2 terbesar, dan kombinasi BF, MI, dan MO terkecil yaitu model ke-30 dengan jumlah BF sebanyak 24, jumlah MI sebanyak 2, dan jumlah MO sebanyak 1 dengan GCV sebesar **0,12011**.

Model terbaik dari kombinasi BF = 24, MI = 2 dan MO = 1 adalah sebagai berikut:

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

Dengan

$$\hat{f}(x) = 0,067 + 0,133BF1 + 0,190BF2 - 0,372BF3 + 0,798BF4 - 1,175BF5 + 0,842BF6 - 0,084BF7 - 0,003BF8 + 0,011BF9 - 0,296BF10 + 0,016BF11 + 0,023BF12$$

Dengan

$$Intercept = 0,067$$

$$BF1 = X_4$$

$$BF2 = X_6$$

$$BF3 = \max(0, X_1 - 17)$$

$$BF4 = \max(0, X_1 - 19)$$

$$BF5 = \max(1, X_1 - 21)$$

$$BF6 = \max(0, X_1 - 22)$$

$$BF7 = \max(0, X_1 - 25)$$

$$BF8 = \max(0, 49 - X_1)$$

$$BF9 = \max(0, X_1 - 49)$$

$$BF10 = \max(0, X_2 * X_4)$$

$$BF11 = \max(0, X_1 - 49) * X_2$$

$$BF12 = \max(0, 49 - X_1) * X_3$$

Dari model terbaik yang dihasilkan, dapat disimpulkan bahwa variabel Usia (X_1), Kesehatan (X_2), Janiman Sosial (X_3), Jenis Kelamin (X_4), Status Kepala Rumah Tangga (X_6) berpengaruh terhadap Partisipasi Tenaga Kerja. Adapun interpretasi dari model yang didapat adalah sebagai berikut:

1. $BF1 = X_4$
Artinya, koefisien dari BF1 akan bermakna jika jenis kelamin (X_4) adalah laki-laki. Partisipasi tenaga kerja yang berjenis kelamin laki-laki cenderung memiliki partisipasi tenaga kerja sebesar 1,143 kali dibandingkan dengan partisipasi tenaga kerja berjenis kelamin perempuan.
2. $BF2 = X_6$
Artinya, koefisien dari BF2 akan bermakna jika status dalam Rumah Tangga (X_6) adalah Kepala Rumah Tangga. Partisipasi tenaga kerja yang berstatus Kepala Rumah Tangga cenderung memiliki partisipasi tenaga kerja sebesar 1,209 kali dibandingkan dengan

partisipasi tenaga kerja berstatus bukan Kepala Rumah Tangga.

3. $BF3 = \max(0, X_1 - 17)$
Artinya, koefisien dari BF3 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 17 tahun. Partisipasi tenaga kerja yang berusia lebih dari 17 tahun cenderung memiliki partisipasi tenaga kerja sebesar 0,689 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 17 tahun.
4. $BF4 = \max(0, X_1 - 19)$
Artinya, koefisien dari BF4 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 19 tahun. Partisipasi tenaga kerja yang berusia lebih dari 19 tahun cenderung memiliki partisipasi tenaga kerja sebesar 2,220 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 19 tahun.
5. $BF5 = \max(1, X_1 - 21)$
Artinya, koefisien dari BF5 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 21 tahun. Partisipasi tenaga kerja yang berusia lebih dari 21 tahun cenderung memiliki partisipasi tenaga kerja sebesar 0,309 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 21 tahun.
6. $BF6 = \max(0, X_1 - 22)$
Artinya, koefisien dari BF6 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 22 tahun. Partisipasi tenaga kerja yang berusia lebih dari 22 tahun cenderung memiliki partisipasi tenaga kerja sebesar 2,320 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 22 tahun dan berjenis kelamin perempuan.
7. $BF7 = \max(0, X_1 - 25)$
Artinya, koefisien dari BF7 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 25 tahun. Partisipasi tenaga kerja yang berusia lebih dari 25 tahun cenderung memiliki partisipasi tenaga kerja sebesar 0,920 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 25 tahun.
8. $BF8 = \max(0, 49 - X_1)$
Artinya, koefisien dari BF8 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih kecil dari 49 tahun. Partisipasi tenaga kerja

yang berusia kurang dari 49 tahun cenderung memiliki partisipasi tenaga kerja sebesar 0,997 kali dibandingkan dengan partisipasi tenaga kerja berusia lebih dari 49 tahun.

9. $BF9 = \max(0, X_1 - 49)$
Artinya, koefisien dari BF9 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 49 tahun. Partisipasi tenaga kerja yang berusia lebih dari 49 tahun cenderung memiliki partisipasi tenaga kerja sebesar 1,011 kali dibandingkan dengan partisipasi tenaga kerja berusia kurang dari 49 tahun.
10. $BF10 = \max(0, X_2 * X_4)$
Artinya, koefisien dari BF10 akan bermakna jika partisipasi tenaga kerja tidak sehat (X_2) dan jenis kelamin laki-laki (X_4). Partisipasi tenaga kerja yang tidak sehat dan laki-laki cenderung memiliki partisipasi tenaga kerja sebesar 0,744 kali dibandingkan dengan partisipasi tenaga kerja yang sehat dan perempuan.
11. $BF11 = \max(0, X_1 - 49) * X_2$
Artinya, koefisien dari BF11 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih besar dari 49 tahun dan partisipasi tenaga kerja tidak sehat. Partisipasi tenaga kerja yang berusia lebih dari 49 tahun dan tidak sehat cenderung memiliki partisipasi tenaga kerja sebesar 1,016 kali dibandingkan dengan partisipasi tenaga kerja yang berusia kurang dari 60 tahun dan tidak sehat.
12. $BF12 = \max(0, 49 - X_1) * X_3$
Artinya, koefisien dari BF12 akan bermakna jika usia (X_1) partisipasi tenaga kerja lebih kecil dari 49 tahun dan tidak memiliki jaminan sosial (X_3). Partisipasi tenaga kerja yang berusia kecil dari 49 tahun dan tidak memiliki jaminan sosial cenderung memiliki partisipasi tenaga kerja sebesar 1,024 kali dibandingkan dengan partisipasi tenaga kerja yang berusia lebih dari 49 tahun dan memiliki jaminan sosial.

Selanjutnya ditunjukkan tingkat kepentingan variabel-variabel prediktor dalam mengklasifikasikan partisipasi tenaga kerja menggunakan metode MARS.

Tabel 4.5 Tingkat Kepentingan Variabel Prediktor

| Variabel | Tingkat Kepentingan |
|---------------------------|---------------------|
| Usia (X_1) | 100,00 |
| Jaminan Sosial (X_3) | 100,00 |
| Status dalam RT (X_6) | 49,2 |
| Jenis Kelamin (X_4) | 42,9 |

Tabel 4.5 menunjukkan tingkat kepentingan variabel prediktor pada klasifikasi model MARS. Berdasarkan hasil analisis yang telah dilakukan, maka didapatkan lima variabel prediktor yang berpengaruh terhadap variabel respon, yaitu Usia (X₁), Jaminan Sosial (X₃), Status Kepala RT (X₆), Jenis Kelamin(X₄), Kesehatan(X₂). Informasi lain yang diperoleh bahwa variabel Usia dan Jaminan Sosial yang mempunyai pengaruh dominan terhadap Partisipasi Tenaga Kerja sebesar 100%. Sedangkan variabel Kesehatan mempunyai pengaruh paling kecil sebesar 32,4%.

4. Klasifikasi Regresi Logistik Biner, MARS dan Bagging MARS

Berikut ini akan disajikan klasifikasi partisipasi tenaga kerja dengan metode Regresi Logistik Biner, metode MARS dan metode Bagging MARS. Dari hasil klasifikasi akan dihitung nilai APER dan TAR. Berikut klasifikasi masing-masing metode.

a. Klasifikasi Regresi Logistik Biner

Persentase ketepatan klasifikasi adalah rasio antara jumlah observasi-observasi yang diklasifikasikan secara tepat oleh model dengan jumlah seluruh observasi. Berikut ini merupakan hasil ketepatan klasifikasi dari metode regresi logistik biner.

Tabel 4.6 Ketepatan Klasifikasi Logistik Biner

| Hasil Observasi | Hasil Prediksi | | Total |
|-----------------|----------------|---------------|-------|
| | Bekerja | Tidak Bekerja | |
| Bekerja | 82 | 27 | 109 |
| Tidak Bekerja | 26 | 33 | 59 |
| Total | 108 | 60 | 168 |

Berdasarkan Tabel 4.6 di atas menunjukkan bahwa dari 168 partisipasi tenaga kerja terdapat 109 orang yang bekerja dan yang diklasifikasikan secara tepat sebanyak 82 orang. Terlihat juga bahwa dari 59 orang yang tidak bekerja terdapat 33 orang yang diklasifikasikan secara tepat. Berikut akan di hitung nilai APER dan TAR untuk hasil klasifikasi dengan regresi logistik biner.

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \cdot 100\% = \frac{26 + 27}{168} \cdot 100\% = 31,55\%$$

$$TAR(\%) = 1 - APER = 1 - 31,55\% = 68,45\%$$

Berdasarkan perhitungan di atas dapat diketahui bahwa model logistik biner tepat mengklasifikasikan partisipasi tenaga kerja sebesar 68,45%, sedangkan 31,55% diklasifikasikan secara tidak tepat.

b. Klasifikasi MARS

Persentase ketepatan klasifikasi adalah rasio antara jumlah observasi-observasi yang diklasifikasikan secara tepat oleh model dengan jumlah seluruh observasi. Berikut ini merupakan hasil ketepatan klasifikasi dari metode MARS.

Tabel 4.7 Ketepatan Klasifikasi Mars

| Hasil Observasi | Hasil Prediksi | | Total |
|-----------------|----------------|---------------|-------|
| | Bekerja | Tidak Bekerja | |
| Bekerja | 102 | 7 | 109 |
| Tidak Bekerja | 22 | 37 | 59 |
| Total | 124 | 44 | 168 |

Berdasarkan Tabel 4.7 di atas menunjukkan bahwa dari 168 partisipasi tenaga kerja terdapat 109 orang yang bekerja dan yang diklasifikasikan secara tepat sebanyak 102 orang. Terlihat juga bahwa dari 59 orang yang tidak bekerja terdapat 37 orang yang diklasifikasikan secara tepat. Berikut akan di hitung nilai APER dan TAR untuk hasil klasifikasi dengan MARS.

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \cdot 100\% = \frac{22 + 7}{168} \cdot 100\% = 17,26\%$$

$$TAR(\%) = 1 - APER = 1 - 17,26\% = 82,74\%$$

Berdasarkan perhitungan di atas dapat diketahui bahwa model MARS tepat mengklasifikasikan partisipasi tenaga kerja sebesar 82,74%, sedangkan 17,26% diklasifikasikan secara tidak tepat.

c. Klasifikasi Bagging MARS

Metode Bagging MARS dilakukan dengan replikasi *bootstrap* pada data partisipasi tenaga kerja sebanyak 50, 100, 150, 200, dan 500 kali, sehingga didapatkan hasil prediksi pada variabel respon dan dapat dilakukan klasifikasi untuk mengetahui ketepatan klasifikasi partisipasi tenaga kerja menggunakan metode Bagging MARS. Berikut hasil ketepatan klasifikasi pada data partisipasi tenaga kerja dengan berbagai kombinasi replikasi.

Tabel 4.8 Ketepatan Klasifikasi Bagging MARS

| Replikasi | TAR | APER |
|------------|---------------|---------------|
| 50 | 80,95% | 19,05% |
| 100 | 80,95% | 19,05% |
| 150 | 82,14% | 17,86% |
| 200 | 82,74% | 17,26% |
| 500 | 83,33% | 16,67% |

Hasil ketepatan klasifikasi metode Bagging MARS yang ditunjukkan pada Tabel 4.8 menunjukkan bahwa dengan replikasi sejumlah 500 kali menghasilkan nilai ketepatan klasifikasi tertinggi sebesar 83,33% dan kesalahan klasifikasi terendah dengan nilai 16,67%. Model yang

digunakan untuk mengklasifikasikan Partisipasi Tenaga Kerja pada Kabupaten Tanah Datar, Sumatera Barat adalah model MARS dengan BF=24, MI=2, dan MO=1.

5. Perbandingan Metode Logistik Biner, MARS, dan Bagging MARS

Untuk mendapatkan metode terbaik yang sesuai digunakan untuk mengklasifikasi partisipasi tenaga kerja antara metode Regresi Logistik Biner, metode MARS dan metode Bagging MARS dapat dilakukan dengan cara membandingkan besarnya nilai ketepatan klasifikasi atau akurasi dari ketiga metode tersebut dengan hasil ditunjukkan sebagai berikut:

Tabel 4.9 Perbandingan Ketepatan Klasifikasi

| Metode | TAR | APER |
|------------------------|---------------|---------------|
| Regresi Logistik Biner | 64,15% | 35,85% |
| MARS | 82,74% | 17,26% |
| Bagging MARS | 83,33% | 16,67% |

Tabel 4.9 menunjukkan perbandingan hasil ketepatan klasifikasi data Partisipasi Tenaga Kerja menggunakan metode Regresi Logistik Biner, MARS dan Bagging MARS didapatkan hasil bahwa metode Bagging MARS lebih tepat digunakan untuk mengklasifikasikan Partisipasi Tenaga Kerja, karena memiliki nilai ketepatan klasifikasi yang lebih besar dibandingkan metode Regresi Logistik Biner dan MARS dengan nilai sebesar 83,33%.

KESIMPULAN

- Jumlah Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat menunjukkan bahwa 64,88% atau sejumlah 109 orang yang bekerja dan 35,12% atau sejumlah 59 orang yang tidak bekerja.
- Dengan menggunakan analisis regresi logistik biner didapatkan model sebagai berikut.

$$g(x) = -1,6227 - 0,6285X_5 + 1,8244X_6$$
Sedangkan model yang diperoleh menggunakan analisis MARS adalah sebagai berikut.

$$\hat{f}(x) = 0,067 + 0,133BF1 + 0,190BF2 - 0,372BF3 + 0,798BF4 - 1,175BF5 + 0,842BF6 - 0,084BF7 - 0,003BF8 + 0,011BF9 - 0,296BF10 + 0,016BF11 + 0,023BF12$$
- Pada analisis regresi logistik biner, variabel yang berpengaruh terhadap Partisipasi Tenaga Kerja adalah variabel Pendidikan (X_5) dan Status dalam RT (X_6). Sedangkan

pada analisis MARS, variabel yang berpengaruh terhadap Partisipasi Tenaga Kerja adalah variabel Usia (X_1), Kesehatan (X_2), Janiman Sosial (X_3), Jenis Kelamin (X_4) dan Status dalam Rumah Tangga (X_6).

- Ketepatan klasifikasi menggunakan analisis regresi logistik biner adalah sebesar 68,45%, ketepatan klasifikasi menggunakan analisis MARS adalah sebesar 82,74% dan ketetapan klasifikasi menggunakan analisis Bagging MARS adalah sebesar 83,33% dengan 500 replikasi. Hal ini menunjukkan bahwa metode Bagging MARS lebih cocok digunakan pada kasus Faktor-Faktor Yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar, Sumatera Barat dibandingkan dengan metode Regresi Logistik Biner dan metode MARS.

DAFTAR PUSTAKA

- Agresti, Allan. 2002. *Categorical Data Analysis*. New York: John Wiley and Sons.
- Arleina, Octiva D. dan Bambang W Otok. 2014. Bootstrap Aggregating Multivariate Adaptive Regression Spline (Bagging MARS) untuk Mengklasifikasikan Rumah Tangga Miskin di Kabupaten Jombang. *Jurnal Sains dan Seni Pomits* 3(2): 163-170.
- Binadari, Ratih., Y. Wilandari., dan Suparti. 2015. Perbandingan Metode Regresi Logistik Biner Dan Multivariate Adaptive Regression Spline (MARS) Pada Peminatan Jurusan sSMA. *Jurnal GAUSSIAN* 4(4): 987-996.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1-67.
- Hasnita, F.M. Afendi., dan A. Fitrianto. 2020. Perbandingan Beberapa Metode Klasifikasi Dalam Memprediksi Interaksi Farmakodinamik. *Indonesian Journal of Statistics and Its Applications* 4(1): 11-21.
- Hosmer and Lemeshow, W. 1989. *Applied Logistic Regression*. Canada: A Wiley Interscience Publication.
- Montgomery, Douglas, Peck, Elizabeth, dan Vining, Geoffrey. (2006). *Introduction to Linear Regression Analysis, Fourth Edition*, USA: Wiley-Interscience Publication.
- Nisa', S. S., dan 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, dan Haryatmi, S. (2006). Bootstrap dalam MARS untuk Klasifikasi Perbankan. *Inferensi Jurnal Statistik*.

Rahmaniah, M.N., Y.Novia., dan I.Purnamasari. 2016. Bootstrap Aggregating Multivariate Adaptive Regression Splines (Studi Kasus : Identifikasi Komponen Penciri Akreditasi Sekolah/Madrasah Pada Tingkat SD/MI di Provinsi Kalimantan Timur Tahun 2015). *Jurnal EKSPONENSIAL* 7(2): 163-170.

Ramandhani, Ridha., Sudarno., dan D.Safitri. 2017. Metode Bootstrap Aggregating Regresi Logistik Biner Untuk Ketepatan Klasifikasi Kesejahteraan Rumah Tangga Di Kota Pati. *Jurnal GAUSSIAN* 6(1): 121-130.

