

# Analisis Regresi Logistik Multinomial dengan Metode Bayes untuk Identifikasi Faktor-Faktor Terjadinya Diabetes Melitus

Sartika Sari Dewi<sup>1</sup>, Resmawan<sup>2</sup>, La Ode Nashar<sup>3</sup>

<sup>1,2,3</sup>Program Studi Statistika, Universitas Negeri Gorontalo, Indonesia  
e-mail: <sup>1</sup>laode.nashar@ung.ac.id

**Abstrak.** Diabetes Melitus merupakan salah satu penyebab kematian terbesar di dunia dimana pada tahun 2020 Indonesia dinyatakan berada pada peringkat ke-7 di dunia dengan 10,7 juta penderita diabetes. Provinsi Gorontalo termasuk kedalam 5 besar penderita diabetes melitus. Terdapat beberapa tipe diabetes melitus yang umumnya terdiri dari DM tipe 1, DM tipe 2, dan DM tipe lainnya yang dapat disebabkan oleh dua faktor yaitu faktor yang tidak bisa diubah dan yang bisa diubah. Model regresi logistik multinomial digunakan untuk meneliti faktor tersebut karena variabel dependen memiliki lebih dari 2 kategori. Untuk mengestimasi parameter model regresi logistik multinomial digunakan metode Bayes. Metode Bayes merupakan metode estimasi parameter yang menghubungkan antara distribusi *prior* dengan fungsi *likelihood* sehingga menghasilkan distribusi *posterior*. Penyelesaian metode Bayes menerapkan simulasi *Markov Chain Monte Carlo* (MCMC) dengan algoritma *Gibbs Sampler*. Data yang digunakan adalah penderita diabetes melitus di Rumah Sakit Toto Kabupaten Bone Bolango Tahun 2021 dengan variabel dependen DM tipe 1, DM tipe 2, dan DM tipe lain. Variabel Independen terdiri dari usia, jenis kelamin, tingkat pendidikan, pekerjaan, dan hipertensi. Hasil pemodelan menunjukkan bahwa variabel yang berpengaruh secara signifikan terhadap penyebab terjadinya diabetes melitus adalah Tingkat Pendidikan Menengah dan Hipertensi. Berdasarkan model yang didapatkan menghasilkan kesalahan klasifikasi sebesar 0,1885% dengan nilai ketepatan klasifikasi sebesar 99,8115% sehingga dapat dikatakan bahwa hasil observasi sudah sesuai dengan hasil prediksi.

**Kata kunci:** Diabetes melitus, Regresi Logistik Multinomial, Metode bayes, MCMC

**Abstract.** *Diabetes Mellitus is one of the biggest causes of death in the world where in 2020 Indonesia is ranked 7th in the world with 10.7 million people with diabetes. Gorontalo Province is included in the top 5 people with diabetes mellitus. There are several types of diabetes mellitus which generally consist of type 1 DM, type 2 DM, and other types of DM which can be caused by two factors, namely factors that cannot be changed and those that can be changed. The multinomial logistic regression model is used to examine these factors because the dependent variable has more than 2 categories. To estimate the parameters of the multinomial logistic regression model, the Bayes method is used. The Bayes method is a parameter estimation method that connects the prior distribution with the likelihood function to produce a posterior distribution. The completion of the Bayes method applies Markov Chain Monte Carlo (MCMC) simulation with the Gibbs Sampler algorithm. The data used are diabetes mellitus sufferers at Toto Hospital, Bone Bolango Regency in 2021 with the dependent variable type 1 DM, type 2 DM, and other types of DM. Independent Variables consist of Age, Gender, Education Level, Occupation, and Hypertension. The modeling results show that the variables that significantly influence the cause of diabetes mellitus are the level of secondary education and hypertension. Based on the model obtained, it resulted in a classification error of 0.1885% with a classification accuracy value of 99.8115%. So it can be said that the observation results are in accordance with the prediction results.*

**Keywords:** *Diabetes mellitus, Multinomial Logistic Regression, Bayes method, MCMC*

## I. PENDAHULUAN

Diabetes melitus diartikan sebagai kelainan yang terjadi pada pankreas dimana insulin (hormon yang memproduksi kadar gula darah) itu tidak bekerja secara efektif yang menyebabkan terjadinya kenaikan kadar gula darah atau biasa disebut *Hiperglikemia* [1]. IDF (*International Diabetes Federation*) memprediksi sekurang-kurangnya pada tahun 2019 terdeteksi sebanyak 463 juta orang di dunia dengan prevalensi sekitar 9,3% menderita diabetes untuk

umur 20-79 tahun.

Infodatin Kementerian kesehatan republik indonesia (2020) menyatakan bahwa Indonesia merupakan peringkat ke-7 di dunia dengan 10,7 juta penderita diabetes. Adapun 5 provinsi di Indonesia yang termasuk penderita terbanyak antara lain DKI Jakarta, Papua Barat, Gorontalo, Banten dan Riau yang mencapai prevalensi sebanyak 0,9% [2]. Hal ini berarti memperlihatkan bahwa provinsi Gorontalo perlu penanganan serius terhadap masalah ini dimana tingkat

prevalensinya yang begitu tinggi. Prevalensi diabetes di Provinsi Gorontalo pada setiap kabupaten/kota yaitu (34,05%) Kab. Gorontalo, (25,3%) Kota Gorontalo, (12,1%) Kab. Boalemo, (11,95%) Kab. Gorontalo Utara, (10,5%) Kab. Bone Bolango, dan (6,15%) Kab. Pohuwato [3]. Pada tahun 2015 DINKES Bone Bolango menemukan beberapa kasus diabetes melitus yaitu sebanyak 251 kasus di Kec. Suwawa Induk, 110 kasus di Suwawa Timur, dan 69 kasus di Suwawa Tengah. Oeh karena itu selama tahun 2014 daerah Suwawa tercatat sebagai urutan ke-9 tertinggi dari banyaknya kasus yang didapatkan [4].

Berdasarkan banyaknya penemuan kasus diabetes melitus tersebut, maka diperlukan perhatian serius untuk mengatasi masalah tersebut. Adapun hal yang dapat dilakukan untuk mengurangi peningkatan kasus diabetes melitus tersebut yaitu dengan memperhatikan faktor apa saja yang signifikansi terhadap terjadinya diabetes melitus. Sebagian faktor yang memiliki kemungkinan mempengaruhi terjadinya diabetes melitus yaitu jenis kelamin, usia, hipertensi, pekerjaan, tingkat pendidikan, Indeks massa tubuh, dan lain sebagainya [2]. Salah satu cara agar mengetahui keterkaitan faktor-faktor tersebut terhadap terjadinya diabetes melitus yaitu dengan memanfaatkan analisis regresi logistik. Regresi logistik adalah teknik dalam statistika yang berfungsi untuk menganalisis keterkaitan antar variabel, dimana variabel dependen bersifat kategorik. Variabel dependen yang memiliki dua kategorik disebut regresi logistik biner. Sementara itu, variabel dependen dengan lebih dari dua kategorik disebut regresi logistik multinomial [5].

Pada penelitian ini, digunakan regresi logistik multinomial untuk menyelesaikan permasalahan tersebut dikarenakan variabel dependen mempunyai lebih dari dua kategorik. Umumnya estimasi parameter yang diterapkan dalam regresi logistik multinomial adalah *maximum likelihood estimation*, akan tetapi permasalahannya adalah peluang terjadinya diabetes melitus terhadap penderita tidaklah berasal dari faktor yang sama, selain itu estimasi parameter dalam metode bayes memiliki sifat-sifat statistik yang baik, prediksi data yang hilang dan peramalan data yang akan datang, deskripsi parsimonius data terobservasi, serta kerangka komputasi untuk estimasi model, pemilihan, dan validasi. Oleh sebab itu, lebih tepat jika menggunakan estimasi bayes atau biasa disebut Metode Bayes [6].

Meliza dkk. (2020) menjelaskan bahwa metode bayes adalah proses estimasi parameter dimana informasi awal (*prior*) digabungkan dengan fungsi *likelihood* yang kemudian hasilnya dinyatakan menjadi distribusi *posterior* [7]. Adapun keunggulan dari estimasi bayes dibandingkan MLE yaitu mampu mendapatkan estimasi parameter yang baik jika yang menjadi patokannya adalah ragam dan kuadrat terkecil galat. Pendekatan yang digunakan yaitu *Markov Chain Monte Carlo* melalui algoritma *Gibbs Sampler* diperoleh nilai estimasi parameter dalam metode bayes yang diperlukan untuk pembentukan model [8].

Beberapa studi yang berkaitan dengan analisis regresi logistik dilakukan dalam berbagai kasus oleh Tulong dkk. (2018) yang menggunakan analisis regresi logistik multinomial untuk mengidentifikasi faktor yang mempengaruhi pilhan siswa di Pulau Tauland terhadap perguruan tinggi [9]. Ndangi, dkk [10] mengulas

perbandingan regresi logistik multinomial dengan analisis diskriminan pada kasus pengklasifikasian siswa berdasarkan kemampuan akademiknya. Hasil penelitian menunjukkan bahwa regresi multinomial dapat dilakukan dengan lebih mudah dengan tingkat akurasi yang sama baiknya. Selanjutnya, Diksa dan Fithriasari [11] menggunakan analisis regresi logistik terhadap kasus yang serupa yaitu faktor resiko terjadinya diabetes melitus namun dalam hal ini menggunakan regresi logistik biner. Kedua kasus tersebut memanfaatkan analisis regresi logistik dalam penyelesaian masalah dengan penduga parameter MLE (*maximum likelihood estimation*). Namun dalam sejumlah kasus, penggunaan regresi logistik perlu dilakukan menggunakan penduga parameter bayes seperti kasus Meliza dkk [7] untuk mengetahui faktor yang mempengaruhi terjadinya kanker payudara di RS Dr. Wahidin Sudirohusodo diterapkan analisis regresi logistik biner berdasarkan metode bayes.

Berdasarkan uraian sebelumnya, maka akan dilakukan penelitian terkait faktor penyebab terjadinya diabetes melitus. Dalam penyelesaiannya akan digunakan analisis regresi logistik multinomial karena menggunakan variabel dependen lebih dari dua kategorik dengan adanya beberapa tipe diabetes melitus yaitu DM tipe 1, DM tipe 2, dan DM tipe lainnya. Selain itu, penduga parameter akan menggunakan metode bayes untuk mendapatkan estimator yang baik dalam penyelesaian kasus ini.

## II. LANDASAN TEORI

### 2.1 Regresi Logistik

Regresi logistik didefinisikan sebagai cara pengolahan statistik yang berfungsi untuk mengetahui keterkaitan antara variabel dependen yang memiliki kategorik dengan variabel independen. Regresi logistik terbagi atas dua berdasarkan jumlah kategorik dependennya yaitu regresi logistik biner dengan variabel dependen 2 kategorik dan regresi logistik multinomial dengan variabel dependen lebih dari 2 kategorik [5]. Berikut adalah persamaan umum dari regresi logistik yaitu:

$$g(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p \quad (1)$$

### 2.2 Regresi Logistik Multinomial

Regresi logistik multinomial didefinisikan sebagai suatu proses regresi untuk mengidentifikasi keterkaitan antara variabel dependen dengan lebih dari 2 kategorik terhadap variabel independen [12]. Asumsikan bahwa variabel dependen Y memiliki tiga kategori dengan kode 0,1, dan 2. Fungsi logit  $Y=1$  banding  $Y=0$  merupakan parameter dependen untuk model regresi logistik biner. Sedangkan fungsi logit  $Y=0$  sebagai pembandingan  $Y=1$  dan  $Y=2$  merupakan parameter dependen untuk model regresi logistik multinomial yang berarti mempunyai dua fungsi logit [6].

Probabilitas bersyarat  $P(y = j|x) = \pi_j(x)$ ,  $j = 0, 1, 2$  Sehingga persamaannya :

$$P(Y = 0|x) = \frac{1}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (2)$$

$$P(Y = 1|x) = \frac{e^{g^1(x)}}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (3)$$

$$P(Y = 2|x) = \frac{e^{g^2(x)}}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (4)$$

Adapun bentuk persamaan regresi logistic multinomial jika diubah ke dalam fungsi logit adalah sebagai berikut:

$$g_1(x) = \ln \left[ \frac{P(Y = 1|x)}{P(Y = 0|x)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p \quad (5)$$

$$g_2(x) = \ln \left[ \frac{P(Y = 2|x)}{P(Y = 0|x)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p \quad (6)$$

Fungsi likelihood dibangkitkan dengan membuat tiga variabel biner dengan kode 0 atau 1 untuk memperlihatkan keanggotaan kelompok dari sebuah pengamatan. Kategori variabel dilakukan seperti berikut: apabila  $Y=0$  sehingga  $Y_0=1$ ,  $Y_1=0$ , dan  $Y_2=0$ ; apabila  $Y=1$  sehingga  $Y_0=0$ ,  $Y_1=1$ , dan  $Y_2=0$ ; apabila  $Y=2$  sehingga  $Y_0=0$ ,  $Y_1=0$ , dan  $Y_2=1$ . Oleh karena itu, berapapun nilai  $Y$ , jumlah dari variabel tersebut  $\sum_{j=1}^2 Y_j = 1$ .

### 2.3 Uji Multikolinearitas

Dalam melakukan analisis regresi logistic multinomial diperlukan data yang tidak memiliki masalah multikolinearitas. Uji multikolinearitas didefinisikan sebagai proses untuk mengetahui adanya keterkaitan yang kuat antar variabel independent. Untuk mengidentifikasi masalah multikolinearitas antar variabel bisa dilihat menggunakan nilai *Variance Inflation Factor* (VIF) [12].

### 2.4 Metode Regresi Logistik Bayes

Metode Bayes pada analisis regresi logistic ini memiliki pola yang serupa dengan semua metode bayes. Metode bayes merupakan metode estimasi parameter yang menghubungkan antara distribusi *prior* dengan fungsi *likelihood* sehingga menghasilkan distribusi *posterior* [7]. Adapun persamaan dari metode bayes dapat dituliskan menjadi:

$$p(\beta|y) = \frac{p(y|\beta)p(\beta)}{p(y)} \quad (7)$$

$p(y)$  adalah fungsi densitas mengikut data  $y$ . Karena tidak mengandung skala parameter  $\beta$ , maka skala  $p(y)$  diibaratkan konstan. Oleh karena itu, persamaannya menjadi:

$$p(\beta|y) = l(\beta|y)p(\beta) \quad (8)$$

Berikut adalah penjelasan mengenai fungsi *likelihood*, distribusi *prior*, dan distribusi *posterior* antara lain yaitu:

#### 1. Fungsi Likelihood

Fungsi *likelihood* merupakan fungsi dari  $\beta$  yang disimbolkan sebagai  $l(\beta)$ . Fungsi *likelihood* dalam regresi

logistik multinomial menirukan distribusi multinomial.

#### 2. Distribusi Prior

Distribusi *prior* merupakan informasi awal yang bersifat subyektif berdasarkan keyakinan seseorang dan memberikan informasi mengenai parameter. Distribusi *prior* diperoleh dari pendekatan parameter  $\beta$  terhadap  $p(\beta)$ . Sehingga bisa dituliskan menjadi:

$$\beta \sim p(\beta) \quad (9)$$

Distribusi *prior* terdiri dari beberapa jenis antara lain [6]:

- 1) *Pseudo prior* adalah memberikan nilai prior sesuai hasil *elaborasi* (penggarapan secara tekun dan cermat) dari *frequentist* atau metode klasik (misalnya regresi dengan metode MLE).
- 2) Distribusi *prior* proper dan improper merupakan pemberian nilai apabila setiap variabel random  $t$  memiliki bentuk tetap atau tidak tetap.
- 3) Hal ini terkait dengan penentuan setiap parameter pada model distribusi sebelumnya.
  - Pemberian *prior* pada model berdasarkan informasi yang ada sebelumnya merupakan *Prior* informatif.
  - Pemberian *prior* pada model tidak berdasarkan informasi yang ada sebelumnya merupakan *Prior* non-informatif.
- 4) Berhubungan dengan bentuk distribusi hasil identifikasi pola datanya.
  - Pemberian *prior* pada model dengan mempertimbangkan pola pembentuk fungsi *likelihood*-nya merupakan *Prior* konjugat.
  - Pemberian *prior* pada model tidak mempertimbangkan pola pembentuk fungsi *likelihood*-nya merupakan *Prior* non-konjugat.

#### 3. Distribusi Posterior

Bentuk analitik yang rumit distribusi *posterior* pada estimasi parameter regresi logistic multinomial mengakibatkan dilakukan metode simulasi. Metode simulasi yang diterapkan yaitu MCMC biasanya menggunakan algoritma *gibbs sampler*. Distribusi *posterior*, dinyatakan dengan teorema bayes yaitu:

$$\text{Distribusi Posterior} \propto \text{Likelihood} \times \text{Distribusi Prior}$$

#### 2.5 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo adalah analisis data dengan cara tiruan numerik yang berfungsi dalam menetapkan skala parameter sebuah integrasi analitik yang susah [13]. Fitriani dkk [14] mendefinisikan MCMC adalah suatu cara memperoleh distribusi posterior yang stasioner dengan teknik tiruan stokastik yaitu menirukan sistem kerja Markov Chain dengan metode Monte Carlo. Metode ini memiliki tiga macam yaitu metode Gibbs Sampling, metode Metropolis, dan metode Metropolis-Hasting. Metode yang

peneliti gunakan adalah Gibbs Sampling karena merupakan metode yang paling lumrah digunakan oleh pengguna metode bayes. Untuk mengetahui hasil konvergensi dari hasil estimasi parameter bayes dapat dilihat dari beberapa plot antara lain yaitu:

1. Trace Plot  
Trace plot dapat dikatakan konvergen apabila sebaran plotnya sudah stabil dan membentuk pola pada satu titik.
2. Density Plot  
Density Plot dapat dikatakan konvergen apabila kurva yang dibentuk berdistribusi normal (seperti lonceng).
3. Quantiles Plot  
Quantiles Plot dapat dikatakan konvergen apabila sebaran plotnya telah stabil dan berada dalam credible interval.
4. Autocorrelation Plot  
Autocorrelation Plot dapat dikatakan konvergen apabila pada lag pertama menuju satu dan lag berikutnya menuju nilai nol.

### 2.6 Interval Kredibel dan Ketepatan Klasifikasi

Dalam uji hipotesis menggunakan parameter bayes dijalankan berdasarkan pendekatan interval kredibel 95% pada setiap parameter. Interval kredibel 95% diukur dari kuantil 2,5% (batas bawah) dan kuantil 97,5% (batas atas). Apabila nilai antara batas bawah dan batas atas tidak memuat nilai nol maka parameter tersebut dinyatakan signifikan. Dalam mencari ketepatan klasifikasi model fungsi yang didapatkan maka diperlukan peluang kesalahan. Nilai peluang sampel yang salah diklasifikasikan dapat dilihat dari nilai APER (Apparent Error Rate). Berikut adalah tabel klasifikasi untuk melihat ketepatan klasifikasi yaitu [6]:

**Tabel 1.** Perhitungan Ketepatan Pengklasifikasian

Kategori Aktual	Kategori Prediksi		
	y = 1	y = 2	y = 3
y = 1	n <sub>11</sub>	n <sub>12</sub>	n <sub>13</sub>
y = 2	n <sub>21</sub>	n <sub>22</sub>	n <sub>23</sub>
y = 3	n <sub>31</sub>	n <sub>32</sub>	n <sub>33</sub>

Ketepatan klasifikasi = 1-APER di mana nilai APER didapatkan dari persamaan berikut.

$$APER (\%) = \left( \frac{n_{12} + n_{13} + n_{21} + n_{23} + n_{31} + n_{32}}{n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}} \right) \times 100\% \quad (10)$$

### 2.7 Diabetes Melitus

Diabetes melitus adalah suatu kelainan metabolisme pada insulin dimana terjadi sekresi insulin atau gangguan kerja [15]. Dinas Kesehatan Bone Bolango (2015) menyatakan bahwa telah didapatkan jumlah kasus diabetes melitus sebanyak 251 kasus di Kec. Suwawa Induk, 110 kasus di

Suwawa Timur, dan 69 kasus di Suwawa Tengah [3]. Hal ini pun disampaikan oleh Badan Pusat Statistik Kabupaten Bone Bolango tahun 2015 bahwa “Selama tahun 2014 daerah Suwawa menduduki urutan ke-9 tertinggi dari banyaknya kasus yang tercatat” [4].

Diabetes Melitus (DM) terdiri dari 3 kategorik yaitu DM tipe 1 merupakan kerusakan yang terjadi terhadap sel beta pankreas (reaksi autoimun). Selanjutnya, DM tipe 2 merupakan ketidakmampuan pankreas dalam menghasilkan insulin yang cukup untuk mengganti insulin resistance. Terakhir, DM tipe lain merupakan kelainan genetik fungsi sel beta sehingga menderita hiperglikemia, diabetes yang timbul akibat adanya penyakit yang mempengaruhi kerja atau produksi insulin, penggunaan obat tertentu secara berlebihan dan lain sebagainya [16].

Layaknya penyakit tidak menular lainnya, Diabetes Melitus memiliki faktor risiko atau faktor penyebab yang berkontribusi sehingga terjadilah suatu penyakit. Faktor risiko diabetes ini terbagi menjadi dua antara lain: yang pertama, faktor yang tidak bisa diubah seperti ras, usia, etnik, jenis kelamin, dan lain-lain. Yang kedua, faktor yang bisa diubah seperti hipertensi, obesitas, merokok, kurang aktivitas fisik, dan lain sebagainya. Berdasarkan faktor risiko tersebut maka peneliti memutuskan untuk mengambil beberapa faktor yang akan menjadi faktor-faktor yang mempengaruhi terjadinya diabetes melitus yaitu usia, jenis kelamin, tingkat pendidikan, pekerjaan, dan hipertensi.

## III. METODE

### 3.1 Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Rumah Sakit Toto Kab. Bone Bolango. Populasi dan sampel yang digunakan merupakan pasien diabetes melitus Rumah Sakit Toto Kab. Bone Bolango Tahun 2021 sebanyak 748 jiwa.

### 3.2 Variabel Penelitian

Variabel penelitian terdiri atas dua antara lain:

1. Variabel dependen (y) yang terdiri dari 3 kategorik, yaitu DM tipe 1 diberi kode 0, DM tipe 2 diberi kode 1, DM tipe lainnya diberi kode 2. Variabel dependen yang digunakan merupakan tipe atau klasifikasi dari diabetes melitus.
2. Variabel independen (x) yang terdiri dari 5 kategorik, yaitu usia, jenis kelamin, tingkat pendidikan, pekerjaan, dan hipertensi. Variabel independen yang digunakan merupakan faktor yang kemungkinan mempengaruhi terjadinya diabetes melitus di Rumah Sakit Toto Kab. Bone Bolango.

**Tabel 2.** Variabel Penelitian

Variabel Penelitian	Keterangan	Kategori	Kode
Y	Tipe DM	DM tipe 1	0
		DM tipe 2	1
		DM tipe lain	2
X1	Usia	≤ 45 tahun	0
		> 45 tahun	1
X2	Jenis Kelamin	perempuan	0
		laki-laki	1
X3	Tingkat Pendidikan	pendidikan dasar	0
		pendidikan menengah	1
		pendidikan tinggi	2
X4	Pekerjaan	tidak bekerja	0
		bekerja	1
X5	Hipertensi	tidak ada	0
		ada	1

3.3 Tahapan Penelitian

Adapun tahapan penelitian dalam analisis regresi logistik multinomial dengan metode bayes antara lain sebagai berikut:

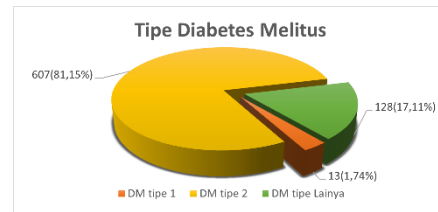
- Menginput data diabetes melitus
- Melakukan analisis deskriptif
- Melakukan uji multikolinearitas  
Apabila nilai VIF  $\geq 10$  artinya terjadi masalah multikolinearitas dan harus di atasi dengan transformasi data. Namun jika VIF  $< 10$  maka tidak terjadi masalah multikolinearitas sehingga proses dapat di lanjutkan ke tahap berikut.
- Melakukan estimasi parameter regresi logistik multinomial dengan MLE tanpa dilakukan uji simultan atau uji parsial menggunakan aplikasi R
- Melakukan estimasi parameter regresi logistik multinomial dengan metode Bayes. Tahapan-tahapan dalam estimasi parameter dengan metode bayes yaitu:
  - Menentukan prior untuk setiap parameter. Prior yang digunakan adalah pseudo prior dimana nilai parameter yang diperoleh dari hasil estimasi parameter menggunakan MLE pada pemodelan regresi logistik multinomial.
  - Melakukan estimasi parameter dengan MCMC melalui Gibbs Sampler sehingga didapatkan hasil akhir yang disebut posterior. Algoritma Gibbs Sampler dilakukan berulang-ulang hingga diperoleh hasil yang konvergen.
  - Menentukan model dan melakukan interpretasi model
- Menghitung peluang untuk masing-masing kategori variabel dependen berdasarkan persamaan (2), (3), dan (4)
- Mencari nilai prediksi untuk variabel dependen berdasarkan Tabel 2.1

- Menghitung persentase ketepatan klasifikasi dengan terlebih dahulu mencari nilai AMPER berdasarkan persamaan (10)
- Mengambil kesimpulan berdasarkan hasil penelitian yang dilakukan.

IV. HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif

Penelitian ini menggunakan lima variabel independen yang diasumsikan sebagai faktor penyebab terjadinya diabetes melitus dengan jumlah sampel sebanyak 748 data. Perbandingan jumlah penderita diabetes melitus untuk DM tipe 1, DM tipe 2, dan DM tipe lainnya disajikan sebagai berikut.



**Gambar 1.** Diagram Lingkaran Variabel Tipe Diabetes Melitus

Berdasarkan **Gambar 1.** dapat dilihat bahwa penderita diabetes melitus paling banyak mengalami DM tipe 2 dengan total 607 pasien yaitu sebesar 81,15%. Penderita diabetes melitus pada DM tipe lainnya dengan total 128 pasien yaitu sebesar 17,11%, dan DM tipe 1 merupakan penderita diabetes melitus paling sedikit yaitu sebesar 1,74% hanya berjumlah 13 pasien.

Deskriptif data untuk variabel independent disajikan dengan memuat kategori setiap variabel, jumlah penderita diabetes disetiap kategori, dan persentase.

**Tabel 3.** Statistika Deskriptif Variabel Independen

Variabel Independen	Kategori	N	%
Usia	≤ 45 tahun	130	17,38
	> 45 tahun	618	82,62
Jenis Kelamin	Perempuan	484	64,71
	Laki-laki	264	35,29
Tingkat Pendidikan	Pendidikan dasar	311	41,58
	Pendidikan menengah	351	46,93
	Pendidikan tinggi	86	11,50
Pekerjaan	Tidak bekerja	430	57,49
	Bekerja	318	42,51
Hipertensi	Tidak ada	425	56,82
	ada	323	43,18

Berdasarkan **Tabel 3.** diperoleh informasi bahwa penderita diabetes melitus paling banyak diderita oleh pasien di atas 45 tahun yaitu sebanyak 618 dengan 82,62% dan paling dominan dialami oleh pasien berjenis kelamin perempuan sebanyak 484 dengan 64,71%. Sementara itu,

untuk penderita diabetes melitus lebih banyak diderita oleh orang yang tingkat pendidikan terakhirnya yaitu pendidikan menengah sebanyak 351 (46,93%), diikuti pada tingkat pendidikan dasar sebanyak 311 (41,58%), dan paling sedikit pada tingkat pendidikan tinggi hanya sebanyak 86 (11,50%). Selanjutnya, untuk penderita diabetes melitus lebih rentan terjadi pada orang yang tidak bekerja yaitu sebanyak 430 dengan 57,49% dan sebanyak 425 dengan 56,82% pasien penderita diabetes melitus tidak memiliki penyakit penyerta hipertensi.

4.2 Uji Multikolinearitas

Pengujian multikolinearitas perlu dilakukan sebelum melakukan estimasi parameter regresi logistik, hal ini bertujuan untuk mengetahui adanya keterkaitan yang kuat antar variabel independent dapat dilihat dari nilai Variance Inflation Factor (VIF).

Hipotesis:

- H<sub>0</sub> : Tidak ada masalah multikolinearitas
- H<sub>1</sub> : Ada masalah multikolinearitas

Dengan kriteria pengujian, apabila nilai VIF < 10 maka terima H<sub>0</sub>, kesimpulannya tidak terdapat masalah multikolinearitas antar variabel. Begitupun sebaliknya. Hasil pengujian dilakukan menggunakan aplikasi R ditampilkan sebagai berikut.

Tabel 4. Uji Multikolinearitas

Variabel	VIF
X1	1,010415
X2	2,169777
X3	1,074849
X4	2,233183
X5	1,012327

Berdasarkan Tabel 4.2 diketahui bahwa setiap variabel independen mempunyai nilai VIF < 10, maka terima H<sub>0</sub>. Kesimpulannya tidak terdapat masalah multikolinearitas antar variabel. Setelah dilakukan uji multikolinearitas, selanjutnya untuk mendapatkan nilai prior dilakukan estimasi parameter model regresi logistik multinomial dengan MLE.

4.3 Estimasi Model Regresi Logistik Multinomial dengan MLE

Tahap ini bertujuan untuk mencari nilai estimasi parameter metode Maximum Likelihood Estimation (MLE) tanpa dilakukan uji simultan atau uji parsial. Dalam tiga kategori variabel dependen menghasilkan dua fungsi logit, salah satu kategori (DM tipe 1) adalah pembanding kategori yang lain yaitu DM tipe 2 dan DM tipe lainnya. Untuk fungsi logit pertama g<sub>1</sub>(x), kategori DM tipe 2 bernilai 1 dan kategori lainnya sebagai pembanding yaitu DM tipe 1 dan DM tipe lainnya bernilai 0. Sedangkan untuk fungsi logit kedua g<sub>2</sub>(x), kategori DM tipe lainnya bernilai 1 dan kategori DM tipe 1 dan DM tipe 2 bernilai 0. Hasil analisis menggunakan aplikasi R dapat dilihat sebagai berikut.

Tabel 5. Estimasi Parameter Regresi Logistik Multinomial dengan MLE

Variabel	Parameter	Estimasi
Konstanta	β <sub>10</sub>	1,63512
Usia	β <sub>11</sub>	0,29403
Jenis Kelamin	β <sub>12</sub>	-0,35059
Tingkat Pendidikan Dasar	β <sub>13</sub> (1)	-0,37488
Tingkat Pendidikan Menengah	β <sub>13</sub> (2)	-0,69358
Pekerjaan	β <sub>14</sub>	0,03729
Hipertensi	β <sub>15</sub>	0,45380
Konstanta	β <sub>20</sub>	-1,82072
Usia	β <sub>21</sub>	-0,19976
Jenis Kelamin	β <sub>22</sub>	0,18186
Tingkat Pendidikan Dasar	β <sub>23</sub> (1)	0,38600
Tingkat Pendidikan Menengah	β <sub>23</sub> (2)	0,70577
Pekerjaan	β <sub>24</sub>	0,08325
Hipertensi	β <sub>25</sub>	-0,47706

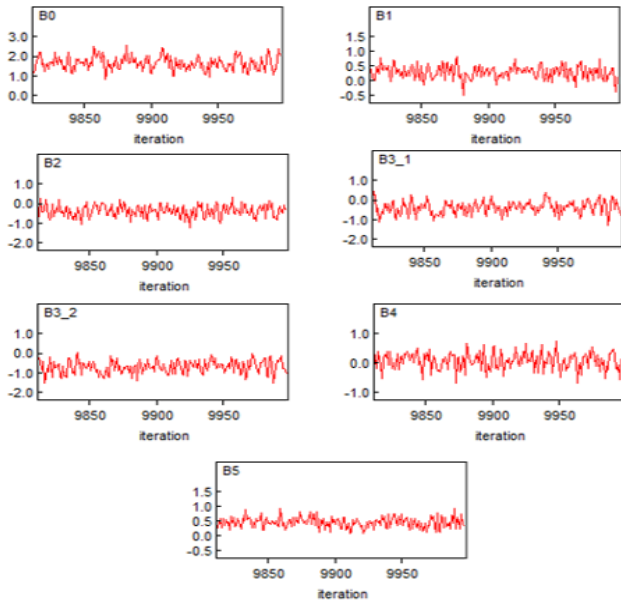
Setelah diperoleh Tabel 5., maka langkah selanjutnya adalah menentukan nilai prior pada estimasi parameter regresi logistik multinomial dengan metode bayes. Dalam hal ini, peneliti menggunakan prior jenis pseudo prior yaitu nilai parameter menggunakan MLE pada pemodelan regresi logistik multinomial.

4.4 Estimasi Model Regresi Logistik Multinomial dengan Bayes

Metode bayes ialah estimasi parameter yang menghubungkan fungsi likelihood dan distribusi prior sehingga menghasilkan distribusi posterior. Distribusi prior untuk β<sub>p</sub> ialah distribusi normal untuk nilai parameter μ<sub>p</sub> merupakan hasil estimasi β<sub>p</sub> yang didapatkan dari frequentist dalam MLE serta σ<sub>p</sub><sup>2</sup> bernilai 1.

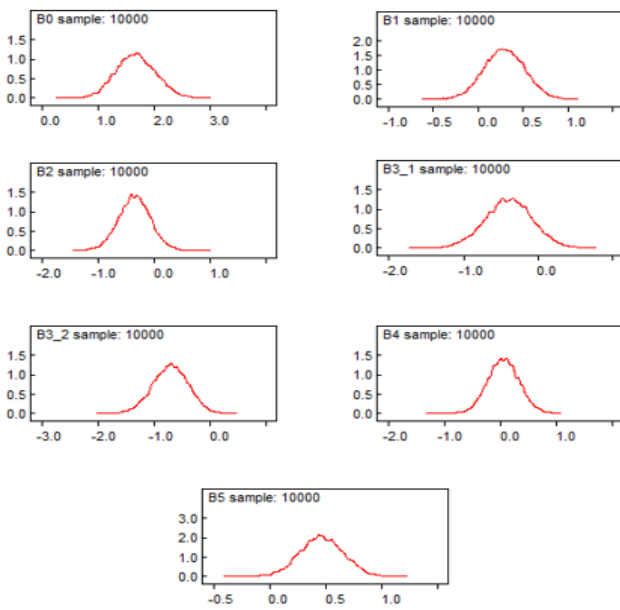
$$\begin{aligned}
 \beta_{10} &\sim dnorm(1,63512,1) & \beta_{20} &\sim dnorm(-1,82072,1) \\
 \beta_{11} &\sim dnorm(0,29403,1) & \beta_{21} &\sim dnorm(-0,19976,1) \\
 \beta_{12} &\sim dnorm(-0,35059,1) & \beta_{22} &\sim dnorm(0,18186,1) \\
 \beta_{13(1)} &\sim dnorm(-0,37488,1) & \beta_{23(1)} &\sim dnorm(0,38600,1) \\
 \beta_{13(2)} &\sim dnorm(-0,69358,1) & \beta_{23(2)} &\sim dnorm(0,70577,1) \\
 \beta_{14} &\sim dnorm(0,03729,1) & \beta_{24} &\sim dnorm(0,08325,1) \\
 \beta_{15} &\sim dnorm(0,45380,1) & \beta_{25} &\sim dnorm(-0,47706,1)
 \end{aligned}$$

Berdasarkan distribusi posterior yang dihasilkan selanjutnya melakukan estimasi dengan simulasi *Markov Chain Monte Carlo (MCMC)* melalui algoritma *gibbs sampler* menggunakan bantuan aplikasi WinBUGS. Kondisi konvergen telah tercapai pada iterasi sebanyak 10.000 dengan thin=10. Kondisi konvergen ditunjukkan MCMC diagnostic plot yang terdiri dari trace plot, density plot, quantiles plot, dan autocorrelation plot. MCMC diagnostic plot untuk setiap parameter ditampilkan pada Gambar 2. s.d. Gambar 5.



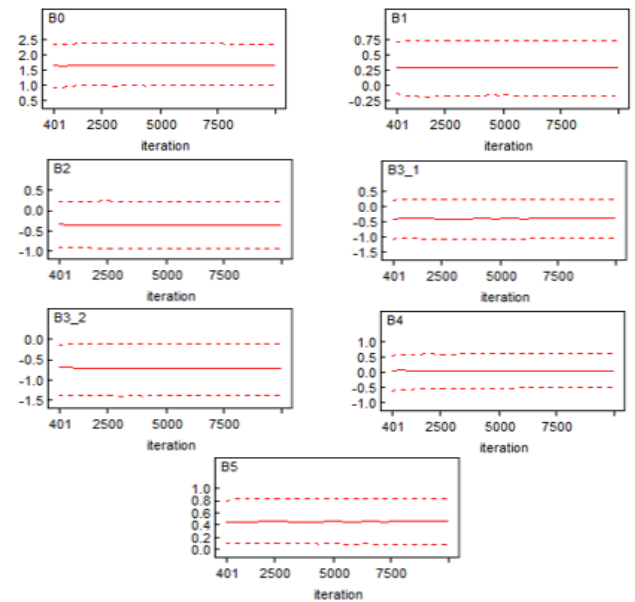
**Gambar 2.** Trace Plot

**Gambar 2.** menunjukkan bahwa parameter yang dihasilkan sudah konvergen dimana sebaran plotnya sudah stabil dan membentuk pola pada satu titik.



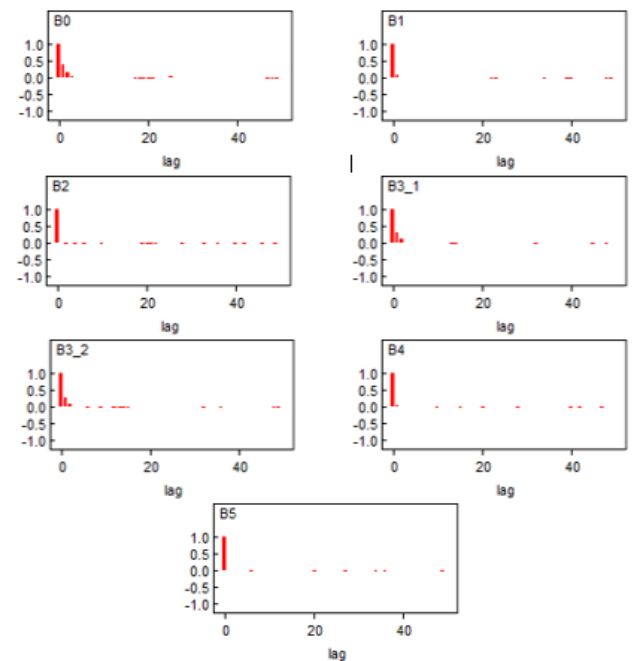
**Gambar 3.** Density Plot

**Gambar 3.** menunjukkan konvergensi pada algoritma sudah tercapai dimana sebaran *posterior* yang terbentuk untuk setiap parameter menggambarkan pola distribusi normal (kurva berbentuk seperti lonceng).



**Gambar 4.** Quantiles Plot

Pada **Gambar 4.** dapat dilihat bahwa parameter sudah berada pada *interval kredibel* dan terlihat dengan garis lurus yang berarti telah stabil.



**Gambar 5.** Autocorrelation Plot

**Gambar 5.** menunjukkan iterasi sudah konvergen dimana nilai autokorelasi pada lag pertama menuju satu dan lag berikutnya menuju nol.

**Tabel 6.** Estimasi Parameter Regresi Logistik Multinomial dengan Bayes

Parameter	Estimasi	Kuantil (2,5%)	Kuantil (97,5%)	Signifikan
$\beta_{10}$	1,671	1,015	2,371	Ya
$\beta_{11}$	0,2893	-0,1591	0,7365	Tidak
$\beta_{12}$	-0,3562	-0,9222	0,2166	Tidak
$\beta_{13(1)}$	-0,395	-1,032	0,2275	Tidak
$\beta_{13(2)}$	-0,7164	-1,361	-0,104	Ya
$\beta_{14}$	0,04794	-0,5031	0,6221	Tidak
$\beta_{15}$	0,4609	0,08527	0,8376	Ya
$\beta_{20}$	-1,867	-2,619	-1,153	Ya
$\beta_{21}$	-0,1925	-0,6611	0,2932	Tidak
$\beta_{22}$	0,185	-0,387	0,7666	Tidak
$\beta_{23(1)}$	0,4107	-0,23	1,082	Tidak
$\beta_{23(2)}$	0,7357	0,1153	1,396	Ya
$\beta_{24}$	0,07809	-0,4955	0,6378	Tidak
$\beta_{25}$	-0,4852	-0,8951	-	Ya
			0,09313	

Pada **Tabel 6.** variabel akan dinyatakan signifikan apabila pada interval kredible antara 2,5% dan 97,5% didalamnya tidak melewati nilai nol. Untuk  $Y_1$  (DM tipe 2), variabel  $X_{3(2)}$  (Tingkat Pendidikan Menengah) dan  $X_5$  (Hipertensi) memiliki nilai signifikan sehingga dapat dikatakan bahwa Tingkat Pendidikan Menengah dan Hipertensi mempengaruhi terjadinya DM tipe 2 pada pasien RS. Toto tahun 2021. Selanjutnya untuk  $Y_2$  (DM tipe lainnya), variabel yang signifikan juga terdiri dari  $X_{3(2)}$  (Tingkat Pendidikan Menengah) dan  $X_5$  (Hipertensi). Hal ini berarti Tingkat Pendidikan Menengah dan Hipertensi mempengaruhi terjadinya DM tipe lainnya pada pasien RS. Toto tahun 2021.

Persamaan Fungsi logit untuk model regresi logistik multinomial bayes pada faktor penyebab diabetes melitus di RS. Toto Kab. Bone Bolango tahun 2021. Fungsi logit dapat dilihat pada persamaan (11) dan persamaan (12).

$$g_1(x) = 1,671 - 0,7164 x_{3(2)} + 0,4609 x_5 \quad (11)$$

$$g_2(x) = -1,867 + 0,7357 x_{3(2)} - 0,4852 x_5 \quad (12)$$

Berdasarkan persamaan (11) dan (12) dapat diketahui bahwa setiap kenaikan 1 faktor Tingkat Pendidikan Menengah akan menurunkan peluang terjadinya DM tipe 2 sebesar 0,7164 kali yang berarti berpengaruh negatif, sedangkan setiap kenaikan 1 faktor variabel Tingkat Pendidikan Menengah akan meningkatkan peluang terjadinya DM tipe lain sebesar 0,7357 kali yang berarti berpengaruh positif. Setiap kenaikan 1 faktor variabel Hipertensi akan menaikkan peluang terjadinya DM tipe 2 sebesar 0,4609 kali yang berarti berpengaruh positif, sedangkan setiap kenaikan 1 faktor variabel Hipertensi akan menurunkan peluang terjadinya DM tipe lain sebesar 0,4852 kali yang berarti berpengaruh negatif.

Selanjutnya mencari nilai ekponensial setiap parameter untuk melakukan interpretasi model. Nilai tersebut ditunjukkan pada **Tabel 7.**

**Tabel 7.** Nilai Eksponensial Setiap Parameter

Variabel	Parameter	Estimasi	Nilai Eksponensial
Konstanta	$\beta_{10}$	1,671	5,31748
Tingkat Pendidikan Menengah	$\beta_{13(2)}$	-0,7164	0,48851
Hipertensi	$\beta_{15}$	0,4609	1,58550
Konstanta	$\beta_{20}$	-1,867	0,15459
Tingkat Pendidikan Menengah	$\beta_{23(2)}$	0,7357	2,08694
Hipertensi	$\beta_{25}$	-0,4852	0,61557

Berdasarkan **Tabel 7.** pada variabel DM tipe 2 bernilai 1 sehingga variabel independen lainnya dianggap konstan, maka setiap penambahan 1 faktor Tingkat Pendidikan Menengah akan meningkatkan peluang seseorang 0,48851 kali menderita DM tipe 2 dan setiap penambahan 1 faktor Hipertensi akan meningkatkan peluang seseorang 1,58550 kali menderita DM tipe 2. Sedangkan pada variabel DM tipe lain bernilai 1 maka variabel independen lainnya dianggap konstan, sehingga setiap penambahan 1 faktor Tingkat Pendidikan Menengah maka akan meningkatkan peluang seseorang 2,08694 kali menderita DM tipe lain, untuk setiap penambahan 1 faktor Hipertensi maka akan meningkatkan peluang seseorang 0,61557 kali menderita DM tipe lain.

Berdasarkan dua fungsi logit pada persamaan (11) dan (12) dan fungsi peluang yang ditunjukkan pada persamaan (2), (3), dan (4) maka diperoleh nilai peluang seseorang menderita penyakit diabetes melitus yang ditunjukkan 10 data awal sebagai berikut.

**Tabel 8.** Nilai Peluang Untuk Setiap Kategori variabel Dependen

No	$P(Y = 0 x)$	$P(Y = 1 x)$	$P(Y = 2 x)$
1	0,104976	0,885035	0,009989
2	0,188071	0,774579	0,03735
3	0,104976	0,885035	0,009989
4	0,15451	0,821605	0,023885
5	0,188071	0,774579	0,03735
6	0,188071	0,774579	0,03735
7	0,255086	0,66262	0,082294
8	0,188071	0,774579	0,03735
9	0,255086	0,66262	0,082294
10	0,188071	0,774579	0,03735
...	...	...	...
748	0,15451	0,821605	0,023885

Berdasarkan hasil perhitungan **Tabel 8.** selanjutnya digunakan untuk menghitung ketepatan klasifikasi sehingga dapat diketahui peluang kesalahan yang didapatkan dari model yang terbentuk. Ketepatan klasifikasi dapat dilihat pada **Tabel 9.**



**Tabel 9.** Perhitungan Ketepatan Pengklasifikasian

Kategori Aktual	Kategori Prediksi		
	y = 1	y = 2	y = 3
y = 1	0	13	0
y = 2	0	607	0
y = 3	0	128	0

Ketepatan klasifikasi = 1-APER di mana nilai APER didapatkan dari persamaan berikut.

$$\begin{aligned}
 \text{APER (\%)} &= \frac{n_{12}+n_{13}+n_{21}+n_{23}+n_{31}+n_{32}}{n_{11}+n_{12}+n_{13}+n_{21}+n_{22}+n_{23}+n_{31}+n_{32}+n_{33}} \times 100\% \\
 &= \frac{13+0+0+0+0+128}{0+13+0+0+607+0+0+128+0} \times 100\% \\
 &= \frac{141}{748} \times 100\% = 0,1885\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Maka ketepatan klasifikasinya } 1-\text{APER} &= 1-0,1885\% \\
 &= 100\%-0,1885\% \\
 &= 99,8115\%
 \end{aligned}$$

Berdasarkan hasil perhitungan nilai APER maka diperoleh kesalahan klasifikasi sebesar 0,1885% dan menghasilkan nilai ketepatan klasifikasi sebesar 99,8115%. Sehingga dapat dikatakan bahwa hasil observasi sudah sesuai dengan hasil prediksi.

#### IV. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan maka diperoleh variabel yang berpengaruh signifikan terhadap terjadinya diabetes melitus di RS. Toto Kab. Bone Bolango Tahun 2021 adalah Tingkat Pendidikan Menengah dan Hipertensi karena nilai parameter didalamnya tidak memuat nilai nol antara kuantil 2,5% dan kuantil 97,5%. Sementara itu berdasarkan nilai APER, kesalahan klasifikasi yang diperoleh sebesar 0,1885% sehingga menghasilkan nilai ketepatan klasifikasi sebesar 99,8115%. Oleh karena itu, dapat dikatakan bahwa hasil observasi sudah sesuai dengan hasil prediksi.

#### REFERENSI

[1] Z. A. Arda, S. Hanapi, Y. Paramata, And A. R. Ngobuto, "Quality Of Life Of Diabetes Mellitus And Determinants In Gorontalo District," *J. Promot. Prev.*, Vol. 3, No. 1, Pp. 14–21, 2020.

[2] Infodatin Kementerian Kesehatan Republik Indonesia, "Tetap Produktif, Cegah Dan Atasi Diabetes Mellitus," *Infodatin Pusat Data Dan Informasi Kementerian Kesehatan Ri*. 2020. [Online]. Available: <https://pusdatin.kemkes.go.id/article/view/20111800001/diabetes-mellitus.html>

[3] F. T. N. Yusuf And Zuhriana K., "Pemanfaatan Tumbuhanobat Sebagai Terapi Pada Penyakit

Diabetes Melitus Di Wilayah Suwawa Kabupaten Bone Bolango," *J. Heal. Dan Sport*, Vol. 14, Pp. 2716–2831, 2017.

[4] Badan Pusat Statistik Kabupaten Bone Bolango, *Suwawa Dalam Angka*. 2015. [Online]. Available: <https://bonebolangokab.bps.go.id/publication/2015/12/31/ee439e473009382702a24bbd/kecamatan-suwawa-dalam-angka-2015.html>

[5] S. R. Dewi, N. Salam, And D. S. Susanti, "Klasifikasi Pemilihan Program Studi Di Fakultas Mipa Universitas Lambung Mangkurat Menggunakan Regresi Logistik Multinomial," *J. Mat. Murni Dan Terap. Epsil.*, Vol. 12, No. 2, Pp. 19–29, 2019, Doi: 10.20527/Epsilon.V12i2.315.

[6] E. T. Cahyani, R. Goejantoro, And Siringoringo, M., "Analisis Regresi Logistik Multinomial Bayes Untuk Pemodelan Minat Peserta Didik Man 2 Samarinda Tahun Ajaran 2018/2019," *J. Eksponensial*, Vol. 13, Pp. 1–8, 2022, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/874%0ahttp://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/download/874/355>

[7] B. D. Meliza, M. N. Bustan, And S. Sudarmin, "Analisis Regresi Logistik Dengan Metode Penduga Bayes Untuk Menentukan Faktor-Faktor Yang Berpengaruh Terhadap Kejadian Kanker Payudara (Studi Kasus: Pasien Kanker Payudara Di Rumah Sakit Dr. Wahidin Sudirohusodo)," *Variansi J. Stat. Its Appl. Teach. Res.*, Vol. 2, No. 2, P. 51, 2020, Doi: 10.35580/Variansiunm12900.

[8] S. Wulandari, F. Yanuar, And H. Yozza, "Model Partisipasi Pemilih Masyarakat Kabupaten Dhamasraya Pada Pemilu 2014 Dengan Menggunakan Metode Regresi Logistik Bayesian," *J. Mat. Unand*, Vol. Vi, No. 1, Pp. 128–133, 2014.

[9] M. M. Tulong, M. Mananohas, And C. E. Mongi, "21456-43748-1-Sm," *J. Mat. Dan Apl.*, Vol. 7, No. 2302–4224, Pp. 90–94, 2018.

[10] W. R. A. Ndangi, R. Resmawan, And I. Djakaria, "Perbandingan Analisis Diskriminan Dan Regresi Logistik Multinomial," *Jambura J. Math.*, Vol. 1, No. 2, Pp. 54–63, 2019, Doi: 10.34312/Jjom.V1i2.2100.

[11] I. G. B. N. Diksa And K. Fithriasari, "Analisis Faktor Resiko Penyebab Diabetes Mellitus Dengan Regresi Logistik Biner," *Inferensi*, Vol. 4, No. 1, P. 69, 2021, Doi: 10.12962/J27213862.V4i1.8480.

[12] N. Setyo Wira Rizki, "Analisis Regresi Logistik Multinomial Pada Pemilihan Alat Kontrasepsi Wanita (Studi Kasus di Puskesmas Sungai Kakap)," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 8, no. 4, pp. 751–758, 2019, doi: 10.26418/bbimst.v8i4.36002.

[13] A. C. Delima, F. Yanuar, And H. Yozza, "Penerapan Metode Regresi Logistik Ordinal

Bayesian Untuk Menentukan Tingkat Partisipasi Politik Masyarakat Kota Padang,” *J. Mat. UNAND*, vol. 8, no. 3, p. 1, 2019, doi: 10.25077/jmu.8.3.1-8.2019.

- [14] V. Aprilia and D. Savitri, “MATH unesa,” *J. Ilm. Mat.*, vol. 7, no. 2, pp. 67–71, 2019.
- [15] F. Milita, S. Handayani, and B. Setiaji, “Kejadian Diabetes Mellitus Tipe II pada Lanjut Usia di Indonesia (Analisis Riskesdas 2018 ),” *J. Kedokt. dan Kesehat.*, vol. 17, no. 1, pp. 9–20, 2021, [Online]. Available: <https://www.bing.com/search?q=prevalensi+diabetes+melitus+di+indonesia+jurnal&cvid=e26a04e174e441dcb70594134d15936a&aqs=edge..69i57.17576j0j9&FORM=ANAB01&PC=ASTS>
- [16] I. B. W. S. H. I. W. P. S. Y. Kardika, “Preanalitik Dan Interpretasi Glukosa Darah Untuk Diagnosa Diabetes Melitus,” *E-Jurnal Med. Udayana*, vol. 2, pp. 1707–1721, 2013, [Online]. Available: <https://ojs.unud.ac.id/index.php/eum/article/view/6698>