

Analisis Faktor Resiko Penyebab Diabetes Mellitus dengan Regresi Logistik Biner

I Gusti Bagus Ngurah Diksa^{1*} and Kartika Fithriasari²

^{1,2}Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember, Surabaya

*Corresponding author: igbn.diksa@bps.go.id

Received: 18 January 2021

Accepted: 27 March 2021

Published: 31 March 2021

ABSTRACT - Diabetes Mellitus menjadi salah satu masalah perawatan kesehatan utama di seluruh dunia. Penyakit gula ini merupakan penyakit berbahaya yang mana mengakibatkan kematian akibat komplikasi yang ditimbulkannya. Banyak faktor yang memengaruhi orang menderita diabetes, beberapa diantaranya yaitu usia, merokok, serum sodium dan platelet dalam badan. Regresi logistik merupakan salah satu alat statistik yang dapat digunakan dalam permodelan klasifikasi tentang ada tidaknya yang mengalami diabetes. Tujuan penelitian ini adalah melihat pengaruh variabel independent usia, merokok, serum sodium dan platelet dalam mengklasifikasikan observasi antara kategori yang tidak mengalami diabetes dan penderita diabetes. Hasil yang didapatkan adalah semua variabel independent signifikan berpengaruh di dalam model dimana semakin meningkatnya umur kecenderungan orang menjadi diabetes semakin tinggi. Selain itu, kegiatan merokok mampu memberikan kecenderungan orang menderita diabetes daripada orang yang tidak merokok. Kemudian semakin bertambahnya serum sodium dalam tubuh maka kecenderungan orang akan tidak menderita diabetes serta untuk bertambahnya platelet memberi kecenderungan sangat kecil orang menderita diabetes. Dalam klasifikasi ini, persentase akurasi klasifikasi sebesar 61,9 persen. Walaupun lebih dari 50 persen namun terjadi misklasifikasi orang yang menderita diabetes sebagai orang tidak mengalami diabetes sebesar 60 persen. Hal itu menyebabkan klasifikasi ini agak beresiko dalam mengelompokkan orang yang diabetes sebagai tujuan penanganan yang lebih cepat.

Keywords diabetes, regresi, logistic, biner, akurasi

I. PENDAHULUAN

Diabetes Mellitus masih menjadi salah satu masalah perawatan kesehatan utama di seluruh dunia di antara populasi yang semakin menua, 422 juta orang dewasa hidup dengan diabetes pada tahun 2014. Data Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa prevalensi diabetes, terutama diabetes tipe 2 yang terus meningkat dalam tiga dekade terakhir, terutama di negara-negara berpenghasilan rendah dan menengah [1]. Diabetes mellitus adalah gangguan metabolisme heterogen yang ditandai dengan adanya hiperglikemia akibat gangguan sekresi insulin [2]. Penyakit gula ini merupakan penyakit berbahaya yang mana mengakibatkan kematian akibat komplikasi yang ditimbulkan dari penyakit diabetes mellitus. Perlunya menyelidiki adanya penyakit diabetes mellitus sejak dini membuat penanganannya lebih cepat. Banyak faktor yang dapat mempengaruhi seseorang dapat menderita penyakit Diabetes Mellitus.

Di negara berkembang penderita diabetes mellitus berusia antara 45-64 tahun dimana usia tergolong masih sangat produktif. Faktor usia memiliki hubungan dengan fisiologi usia tua dimana semakin tua usia seseorang, maka fungsi tubuh juga mengalami penurunan, termasuk kerja hormon insulin sehingga tidak dapat bekerja secara optimal dan menyebabkan tingginya kadar gula darah [3]. Faktor lainnya yaitu kebiasaan merokok sebagai faktor resiko untuk penyakit jantung koroner. Namun, setelah bertahun-tahun, pengumpulan data penelitian menunjukkan bahwa perokok yang merokok dalam waktu yang lama/kronik mempunyai resiko yang lebih tinggi juga untuk terjadinya resistensi insulin [4]. Hiperglikemia meningkatkan osmolalitas serum, mengakibatkan terjadinya pergerakan air keluar dari sel dan kemudian dalam penurunan kadar natrium serum ($[Na]$) dalam tubuh [5]. Selain itu, Hiperglikemia dapat meningkatkan reaktivitas platelet dengan cara menginduksi *non enzymatic glycation* pada permukaan platelet [6]. Pada diabetes, kalium atau sodium sangat berguna untuk meningkatkan kepekaan insulin, sehingga proses pengurusan gula dalam darah berlangsung efektif [7].

Regresi logistik merupakan salah satu tools statistic yang dapat digunakan dalam permodelan klasifikasi tentang ada tidaknya yang mengalami diabetes. Regresi logistik adalah bentuk khusus analisis regresi dengan respon biner dan prediktor yang dapat terdiri dari data kontinu, kategori atau campuran antara keduanya. Analisis ini tidak memerlukan asumsi distribusi multivariate normal atau kesamaan matrik varian kovarian serta dapat juga diterapkan dalam berbagai skala data [8]. Tujuan penelitian ini adalah melihat pengaruh variabel independent usia, merokok, serum sodium dan platelet dalam mengklasifikasikan observasi antara kategori yang tidak mengalami diabetes dan penderita diabetes.

II. MATERI DAN METODE

A. Sumber Data

Data yang digunakan dalam penelitian ini data sekunder clinical_data set yang berjumlah 299 observasi. Variabel yang digunakan terdapat lima variabel yaitu

Tabel 1. Variabel penelitian

Variabel	Keterangan	Range	Jenis Data
X ₁	Platelets	[25.01,..., 850.00]	Kontinu
X ₂	Serum Sodium	[114,..., 148]	Kontinu
X ₃	Kebiasaan Merokok	0 : Tidak 1 : Iya	Kategorik
X ₄	Usia	0 : ≤ 51 1 : > 51 & ≤ 60 2 : > 60 & ≤ 70 3 : > 70	Kategorik
Y	Diabetes	0 : Tidak 1 : Iya	Kategorik

Variabel lainnya tidak digunakan dalam penelitian ini. Untuk mempermudah analisis, variabel usia yang sebelumnya merupakan variabel kontinu diubah menjadi kategorik dengan menggunakan batas batas kuartilnya yaitu pada kuartil pertama dengan umur 51 tahun, kuartil kedua 60 tahun dan kuartil 3 adalah 70 tahun.

Dalam penelitian ini digunakan metode regresi logistik biner. Regresi Logistik menggambarkan dan mengestimasi hubungan antar variabel respon (*y*) yang berupa kategorik dengan variabel prediktor (*x*) yang berpengaruh pada suatu fenomena [9]. Model regresi logistik biner merupakan model regresi logistik yang mana variabel responnya berupa data kualitatif dikotomi yaitu bernilai 0 dan 1. Bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Model regresi logistik biner digunakan jika mengikuti distribusi Bernoulli, yang mana persamaannya sebagai berikut :

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1 - y_i}, y = 0, 1 \tag{1}$$

Dimana :

π_i = peluang kejadian ke-*i*

y_i = peubah acak ke-*i* yang terdiri dari 0 dan 1

Bentuk model regresi logistik biner dengan beberapa variabel prediktor adalah :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \tag{2}$$

Untuk mempermudah menaksir parameter regresi, maka $\pi(x)$ pada persamaan diatas ditransformasikan sehingga menghasilkan bentuk logit regresi logistic, sebagai berikut [10] :

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \tag{3}$$

B. Uji Independensi

Uji independensi digunakan untuk mengetahui hubungan antar dua variabel.

H₀ : Tidak ada hubungan antara variabel dependen dengan independen

H₁ : Ada hubungan antara dua variabel dependen dengan independen

Statistik uji yang digunakan adalah sebagai berikut

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \tag{4}$$

atau

$$G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right) \tag{5}$$

Dengan :

$$e_{ij} = \frac{n_i \cdot n_j}{n..} \tag{6}$$

Keterangan :

n_{ij} : nilai observasi / pengamatan baris ke-i kolom ke-j

e_{ij} : nilai ekspektasi baris ke-i kolom ke-j

Daerah kritis : Tolak H_0 jika G^2 atau $\chi^2_{hitung} > \chi^2_{(1-\alpha)(J-1)}$ atau $P\text{-value} < \alpha$. [11]

C. Pendugaan Parameter

Pada dasarnya metode maximum likelihood Estimation (MLE) memberikan nilai estimasi β untuk memaksimalkan fungsi *likelihood*. Secara sistematis fungsi *likelihood* untuk model regresi logistic biner sebagai berikut [17] :

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \tag{7}$$

Dimana :

y_i = pengamatan pada variabel ke-i

$\pi(x_i)$ = peluang untuk variabel prediktor ke-i

Untuk memudahkan perhitungan maka dilakukan pendekatan *log likelihood*, didefinisikan sebagai :

$$L(\beta) = \sum_{i=1}^n \{ y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \} \tag{8}$$

Untuk mendapatkan nilai penafsiran koefisien regresi logistic ($\hat{\beta}$) dilakukan dengan membuat turunan pertama $L(\beta)$ terhadap β dan disamakan dengan 0.

D. Uji Simultan Model Regresi Logistik

Uji model dilakukan untuk memeriksa peranan variabel prediktor terhadap variabel respon secara serentak atau secara keseluruhan disebut juga uji model *chi square*.

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

H_1 : paling sedikit ada satu parameter $\beta_i \neq 0$, dengan $i = 1, 2, \dots, p$

Statistik uji

$$G = -2 \ln \frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\sum_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)}} \tag{9}$$

Dimana :

n_1 = banyaknya observasi yang berkategori 1

n_0 = banyaknya observasi yang berkategori 0

n = jumlah pengamatan

p = jumlah parameter

Statistik uji G mengikuti distribusi *chi-square*, sehingga untuk memperoleh keputusan dilakukan perbandingan dengan nilai χ^2 tabel, dengan derajat bebas (db) = $k - 1$, k merupakan banyaknya variabel predictor. Kriteria penolakan (tolak H_0) jika nilai $G > \chi^2_{(db,\alpha)}$ atau jika $P\text{-Value} < \alpha$ [12].

E. Uji Parsial

Pengujian parsial digunakan untuk menguji pengaruh setiap β_i secara individual dalam model yang diperoleh.

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$ dengan $j = 1, 2, 3, \dots, p$

Statistik uji Wald (W) :

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \text{ Dan } SE(\hat{\beta}_i) = \sqrt{\sigma^2(\hat{\beta}_i)} \tag{10}$$

Dimana :

$SE(\hat{\beta}_i)$ = dugaan galat baku untuk koefisien β_i

$\hat{\beta}_i$ = nilai dugaan untuk parameter (β_i)
 Daerah penolakan : Tolak H_0 jika $W > Z_{\alpha/2}$ [12].

F. Uji Kesesuaian Model

Pengujian ini dilakukan untuk menguji apakah model yang dihasilkan berdasarkan regresi logistik *multivariat*/serentak sudah layak. Pengujian ini menggunakan statistik uji Hosmer dan Lemeshow dengan hipotesis :
 H_0 : tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model.
 H_1 : terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model.
 Statistik Uji:

$$C = \sum_{k=1}^g \frac{(O_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{11}$$

Daerah Penolakan : Tolak H_0 jika $C > \chi^2_{(g-2, \alpha)}$

Keterangan :

O_k : observasi pada grup ke-k

$\bar{\pi}_k$: rata rata taksiran peluang $\left(\sum_{j=1}^{c_k} \frac{m_j \pi_j}{n_k} \right)$

g : jumlah grup (kombinasi kategori dalam serentak)

n_k : banyaknya observasi pada grup ke - k [13].

G. Uji Asumsi Multikolinieritas dan Linieritas

Uji multikolinieritas bertujuan untuk menguji apakah dalam model regresi terdapat korelasi atau hubungan yang tinggi antar variabel bebas. Model regresi yang baik seharusnya tidak terjadi korelasi tinggi diantara variabel bebas (tidak terjadi multikolinieritas). Untuk mendeteksi ada tidaknya gejala multikolinieritas dapat dilakukan dengan melihat pada nilai *variance inflation factor* (VIF) dimana tidak terjadi kasus multikolinieritas apabila nilai VIF kurang dari 10 [14].

Regresi logistik tidak memerlukan distribusi normal multivariat, tetapi memerlukan pengambilan sampel yang acak, independen, dan linieritas antara X dan logit. Model tersebut kemungkinan besar paling akurat bila distribusinya terpusat namun kurang akurat apabila terdapat data ekstrem. Meskipun seseorang dapat memperkirakan P ($Y = 1$) untuk kombinasi nilai apapun, namun kemungkinan tidak semua kombinasi benar-benar ada dalam populasi tersebut [15].

H. Uji Ketepatan Model

Pada penelitian ini, uji ketepatan model menggunakan *Apparent Error Rate* (APER). APER merupakan fraksi observasi dalam sampel yang salah diklasifikasikan atau misclassified pada fungsi klasifikasi. penjelasannya akan dapat dilihat pada tabel 2.

Tabel 2. *Apparent error rate (APER)*

Data Aktual	Data Prediksi		Total
	Y ₁	Y ₂	
Y ₁	n _{1C}	n _{1M}	n ₁
Y ₂	n _{2M}	n _{2C}	n ₂

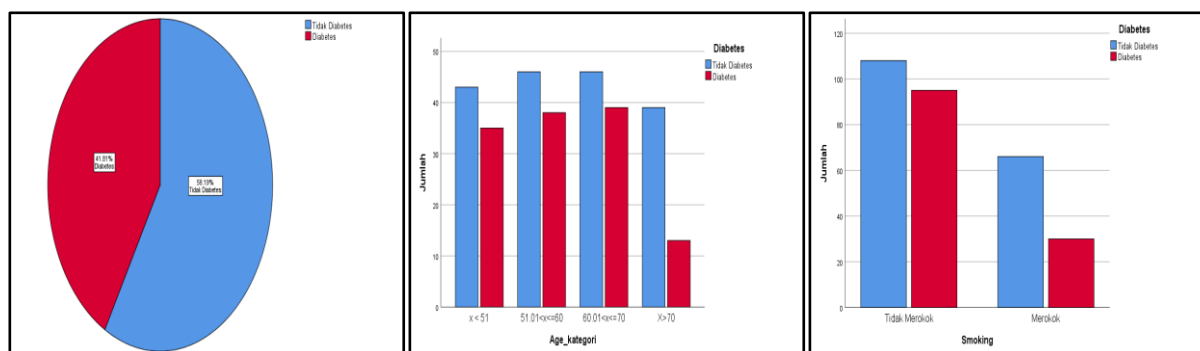
Berdasarkan tabel nilai APER diperoleh berdasarkan persamaan

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{12}$$

III. HASIL DAN PEMBAHASAN

A. Analisis Deskriptif

Berikut disajikan gambar 1 terkait deskripsi dari variabel diabetes tentang persentase orang yang menderita diabetes dan tidak. Kemudian jumlah penderita diabetes dan tidak penderita diabetes disetiap kelompok umur. Dan selanjutnya jumlah penderita diabetes dan bukan penderita diabetes pada kegiatan merokok dan tidak merokok.



Gambar 1 Deskriptif data yang digunakan

Dalam variabel yang digunakan tidak ditemukan adanya *missing value*. Oleh karena itu data dapat digunakan secara lebih lanjut. Berdasarkan gambar 1 dari data 299 observasi 41,81 persen mengalami diabetes mellitus dan 58,19 persen diindikasikan tidak menderita diabetes mellitus. Selain itu di tiap kelompok umur didominasi oleh observasi yang diindikasikan mengalami diabetes. Tidak hanya itu, observasi didominasi oleh observasi yang tidak melakukan kegiatan merokok

B. Uji Independensi

Untuk mengetahui hubungan antara variabel predictor dengan variabel respon dilakukan pengujian independensi dengan $\alpha = 0,10$. Berikut merupakan tabel uji independensi yaitu tabel 3 untuk setiap variabel jenis kelamin dan domisili menggunakan pengujian *chi-square*.

Tabel 3. Uji independensi

Variabel	χ^2_{hitung}	Df	P-Value	Keterangan
Platelets	242,719	175	0,001	Ada Hubungan
Serum Sodium	40,320	26	0,036	Ada Hubungan
Kebiasaan Merokok	6,602	1	0,010	Ada Hubungan
Usia	7,70	3	0,053	Ada Hubungan

Berdasarkan tabel 3 diketahui bahwa semua variabel signifikan. ini berarti ada hubungan antara terjangkitnya diabetes dengan platelets, serum sodium, kebiasaan merokok dan usia. Hal ini ditunjukkan oleh nilai *p-value* yang kurang dari $\alpha = 0,10$

C. Pendugaan Parameter

Adapun persamaan logit yang di dapat adalah

$$\pi(x) = \frac{\exp(5,138 + 0,000002x_1 - 0,055x_2 + 0,711x_{31} + 0,885x_{41} + 1,055x_{42} + 1,062x_{43})}{1 + \exp(5,138 + 0,000002x_1 - 0,055x_2 + 0,711x_{31} + 0,885x_{41} + 1,055x_{42} + 1,062x_{43})} \tag{13}$$

Keterangan

- x_1 : Platelets
- x_2 : Serum Sodium
- x_{31} : Merokok
- x_{41} : Kelompok usia > 51 Tahun dan ≤ 60
- x_{42} : Kelompok usia > 60 Tahun dan ≤ 70
- x_{43} : Kelompok usia > 70

D Uji Simultan

Berikut disajikan hasil uji dari omnibus test terkait uji simultan dari model yang terdapat pada tabel 4.

Tabel 4. Uji serentak (*omnibus tests of model coefficients*)

	Chi-square	df	P-Value
Step	21,153	2	0,000
Block	21,153	2	0,000
Model	21,153	2	0,000

Keterangan : alpha(α) = 10%

Berdasarkan tabel 2 dapat dilihat bahwa nilai *p-value* sebesar 0,000 yang mana nilai tersebut kurang dari alpha (α) sebesar (0,10), sehingga dapat diputuskan Tolak H_0 , sehingga dapat disimpulkan bahwa minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap model.

E. Uji Parsial

Berikut disajikan tabel 5 terkait nilai signifikansi dari tiap variabel independent di dalam model.

Tabel 5. Uji signifikansi parameter secara parsial

Variabel	B	Wald	Df	P-Value
Platelets	0,000002	3,751	1	0,053
Serum Sodium	-0,055	3,637	1	0,057
Merokok	0,711	6,981	1	0,008
Kelompok usia > 51 Tahun dan ≤ 60	0,885	6,968	1	0,028
Kelompok usia > 60 Tahun dan ≤ 70	1,055	7,090	1	0,008
Kelompok usia > 70	1,062	6,981	1	0,008
Constant	5,138	1,762	1	0,184

Tabel 5 menunjukkan bahwa pada tingkat signifikansi 10 % variabel platelets, serum sodium, merokok, usia pada kategori 1, 2 dan 3 berpengaruh signifikan terhadap klasifikasi antara kategori orang menderita penyakit Diabetes Mellitus dan tidak.

F. Uji Kesesuaian Model

Selanjutnya dilakukan uji kesesuaian model yang mana outputnya dapat dilihat dari tabel 6.

Tabel 6. Uji kesesuaian model (*hosmer and lemeshow test*)

Chi-Square	Df	P-Value
8,315	8	0,403

Tabel 5 menunjukkan bahwa nilai *p value* sebesar 0,403 dan itu lebih besar dari α sebesar 0,1 yang berarti gagal tolak H_0 sehingga dapat diputuskan bahwa model sesuai artinya tidak terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model

G. Uji Asumsi Multikolinieritas dan Linieritas

Berdasarkan tabel 7 terlihat bahwa model telah memenuhi asumsi non multikolinieritas dan linieritas yang ditunjukkan dari nilai VIF dan signifikan *p-value* dari asumsi linieritas.

Tabel 7. Uji asumsi multikolinieritas dan linieritas

Variabel	VIF	Keterangan	P- Value Linieritas	Keterangan
Platelets	1,03	Non Multiko	0,3399	Linier
Serum Sodium	1,03	Non Multiko	0,7265	Linier
Merokok (x_{31})	1,01	Non Multiko	-	-
Kelompok usia > 51 Tahun dan ≤ 60	1,50	Non Multiko	-	-
Kelompok usia > 60 Tahun dan ≤ 70	1,50	Non Multiko	-	-
Kelompok usia > 60 Tahun dan ≤ 70	1,31	Non Multiko	-	-
Constant				

H. Interpretasi Model

Berdasarkan hasil uji asumsi dan kelayakan model maka di dapatkan model terbaik dengan nilai odds ratio dari masing masing variabel.

Tabel 8. Nilai odds ratio

Variabel	B	Odd Ratio	P-Value
Platelets	0,000002	1,000	0,053
Serum Sodium	-0,055	0,947	0,057
Merokok	0,711	2,036	0,008
Kelompok usia > 51 Tahun dan ≤ 60	0,885	2,422	0,028
Kelompok usia > 60 Tahun dan ≤ 70	1,055	2,873	0,008
Kelompok usia > 70	1,062	2,893	0,008
Constant	5,138	170,319	0,184

Berdasarkan tabel 8 menunjukkan bahwa nilai odd ratio mayoritas lebih dari satu. Apabila orang itu usianya semakin tua, maka memiliki kecenderungan yang besar untuk orang menderita diabetes yang terlihat dari kelompok umur lebih 70 tahun memiliki kecenderungan orang menderita diabetes adalah 2,893 kali. Selain itu, orang yang merokok akan cenderung menjadi penderita diabetes 2,036 kali dibandingkan orang yang tidak merokok. Hal ini berbeda dengan serum sodium yang mana semakin bertambahnya serum sodium dalam tubuh maka kecenderungan orang untuk menderita diabetes sebesar 0,947 kali. Sedangkan untuk semakin bertambahnya Platelets kecenderungan sangat kecil untuk orang menderita penyakit diabetes.

I. Hasil Ketepatan Model

Setelah dilakukan uji untuk mendapatkan model terbaik maka dilanjutkan menilai hasil klasifikasi yang dihasilkan oleh model yang tergambar pada Tabel 9.

Tabel 9. Ketepatan klasifikasi model

Data Aktual	Data Prediksi		Persentase kebenaran
	Tidak Diabetes	Diabetes	
Tidak Diabetes	135	39	77,6%
Diabetes	75	50	40,00%
Persentase keseluruhan	70,23%	29,77%	61,90%

Tabel 9 menunjukkan bahwa ketepatan orang yang tidak diabetes adalah 77,6 persen dan sisanya terjadi misklasifikasi. Namun terjadi keparahan terhadap misklasifikasi orang yang diabetes yang didiagnosa tidak diabetes yaitu sebesar 60 persen. Kesalahan ini membuat orang untuk lambat dalam menangani penyakit diabetes pada dirinya. Walaupun secara persentase keseluruhan ketepatan klasifikasi adalah 61,90 persen namun kesalahan klasifikasi orang yang diabetes dan menjadi tidak diabetes sangat besar yaitu 60 persen justru mengurangi manfaat dari tujuan penelitian ini.

IV. KESIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan maka didapatkan simpulan sebagai berikut : variabel independen yaitu usia, merokok, serum sodium dan platelets. berpengaruh signifikan terhadap pengklasifikasian orang yang menderita diabetes dan yang tidak menderita diabetes. Semakin bertambah usia seseorang, maka memiliki kecenderungan yang besar untuk menderita diabetes yang terlihat dari kelompok umur lebih 70 tahun memiliki kecenderungan orang menderita diabetes adalah 2,893 kali. Selain itu, orang yang merokok akan cenderung menjadi penderita diabetes 2,036 kali dibandingkan orang yang tidak merokok. Hal ini berbeda dengan serum sodium yang mana semakin bertambahnya serum sodium dalam tubuh maka kecenderungan orang untuk menderita diabetes sebesar 0,947 kali. Sedangkan untuk semakin bertambahnya Platelets kecenderungan sangat kecil untuk orang menderita penyakit diabetes. Berdasarkan perhitungan akurasi, penggunaan model regresi logistic biner ini memberikan akurasi sebesar 61,9 persen. Akan tetapi, terjadi kesalahan besar dalam klasifikasi yang mana penderita diabetes yang diklasifikasi menjadi bukan penderita diabetes sebanyak 60 persen. hal tersebut membuat ketidakefisienan dalam melakukan klasifikasi. Untuk penelitian selanjutnya bisa digunakan *supervised analyze* dalam klasifikasi lainnya yang memiliki tingkat keakuratan yang lebih baik dalam kasus ini. Bila dilihat dari sisi kesehatan dapat disarankan dalam bertambahnya usia maka kita perlu menjaga kesehatan dengan pola hidup sehat dan menghindari diri dari kegiatan merokok.

REFERENSI

[1] K. Tachkov *et al.*, "Life expectancy and survival analysis of patients with diabetes compared to the non diabetic population in Bulgaria," *PLoS One*, vol. 15, no. 5, pp. 1–16, 2020, doi: 10.1371/journal.pone.0232815.
 [2] Punthakee, Zubin, Ronald Goldenberg, and Pamela Katz. "Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome." *Canadian journal of diabetes* 42 (2018): S10-S15.
 [3] B. Kurniawaty, Evi; Yanita, "Faktor-Faktor yang Berhubungan dengan Kejadian Diabetes Melitus Tipe II," *Majority*, vol. 5, no. 2, pp. 27–31, 2016, [Online]. Available: <http://juke.kedokteran.unila.ac.id/index.php/majority/article/view/1073>.

- [4] M. Dwi Ario, "Effect of Nicotine in Cigarette for Type 2 Diabetes Mellitus," *J Major.*, vol. 3, no. 7, pp. 75–80, 2014.
- [5] G. Liamis, "Diabetes mellitus and electrolyte disorders," *World J. Clin. Cases*, vol. 2, no. 10, p. 488, 2014, doi: 10.12998/wjcc.v2.i10.488.
- [6] D. J. Schneider, "Factors contributing to increased platelet reactivity in people with diabetes," *Diabetes Care*, vol. 32, no. 4, pp. 525–527, 2009, doi: 10.2337/dc08-1865.
- [7] R. Nurpalah, "Gambaran Kadar Kalium Pada Penderita Diabetes Melitus Tipe 2," *J. Kesehat. Bakti Tunas Husada J. Ilmu-ilmu Keperawatan, Anal. Kesehat. dan Farm.*, vol. 12, no. 1, p. 214, 2015, doi: 10.36465/jkbth.v12i1.81.
- [8] F. K. Lembang and D. L. Rahakbauw, "Analisis Faktor Risiko Penyebab Diabetes Mellitus di Kota Ambon Menggunakan Model Regresi Logistik," vol. 15, no. 2, pp. 65–71, 2015.
- [9] A. Borucka, "Logistic regression in modeling and assessment of transport services," *Open Eng.*, vol. 10, no. 1, pp. 26–34, 2020, doi: 10.1515/eng-2020-0029.
- [10] Y. A. Tampil, H. Komalig, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado Logistic Regression Analysis To Determine Factors Affecting The Grade Point Average (GPA) Of FM," *d'CARTESIAN*, vol. 6, no. 2, pp. 57–62, 2017.
- [11] K. M. Kotimah and P. S. Wulandari, "Model Regresi Logistik Biner Stratifikasi Pada Partisipasi Ekonomi Perempuan Di Provinsi Jawa Timur," *J. SAINS DAN SENI POMITS Vol. 3, No.1, 2337-3520 (2301-928X Print)*, vol. 3, no. 1, pp. 2337–3520, 2014.
- [12] E. T. (Institusi T. S. N. Pamungkas, "Metode regresi logistik biner pada faktor yang mempengaruhi kesembuhan pasien penderita demam berdarah dengue di RSUD Dr. Iskak Kabupaten Tulungagung," 2017, [Online]. Available: <http://repository.its.ac.id/42259/1/1314030088-Non-Degree.pdf>.
- [13] D. W. Hosmer and S. Lemeshow, "1_3kOQSTg.pdf." pp. 1–375, 2000.
- [14] J. Matematika, F. Matematika, D. A. N. Ilmu, and P. Alam, "PEMODELAN STATUS BEKERJA DI KOTA SEMARANG," 2017.
- [15] D. Berger, "Introduction to Binary Logistic Regression and Propensity Score Analysis," *ResearchGate*, no. October, pp. 1–30, 2017.
- [16] K. M. Kotimah and P. S. Wulandari, "Model Regresi Logistik Biner Stratifikasi Pada Partisipasi Ekonomi Perempuan Di Provinsi Jawa Timur," *J. SAINS DAN SENI POMITS Vol. 3, No.1, 2337-3520 (2301-928X Print)*, vol. 3, no. 1, pp. 2337–3520, 2014.
- [17] Hosmer, D.W., dan S. Lemeshow. 2000. *Applied Logistic Regression*. Edisi ke-2. John Wiley and Sons Inc, Canada.