

Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik

Ermawati

Prodi Matematika FST, Universitas Islam Negeri Alauddin Makassar, ermawati@uin-alauddin.ac.id

Risnawati Ibas

Prodi Matematika FST, Universitas Islam Negeri Alauddin Makassar, risnawati.ibnas@uin-alauddin.ac.id

Baso Agung Kurniawan

Mahasiswa prodi Matematika FST, Universitas Islam Negeri Alauddin Makassar, agung7102001@gmail.com

ABSTRAK. Penelitian ini membahas tentang klasifikasi penderita anemia pada remaja putri dengan berbagai faktor yang mempengaruhinya. Penelitian ini bertujuan untuk mengetahui model klasifikasi penyakit anemia. Penelitian ini merupakan penelitian terapan dimana menerapkan teknik data *mining* yaitu metode regresi logistic untuk mengklasifikasi faktor-faktor penyebab anemia pada remaja wanita. Atribut yang digunakan yaitu faktor-faktor yang mempengaruhi anemia diantaranya *ferritin serum*, STfR, dan riwayat penyakit kronis. Hasil penelitian menunjukkan bahwa terdapat dua variabel yang berpengaruh secara signifikan dalam mengklasifikasi anemia yaitu *ferritin serum*, dan STfR dengan tingkat akurasi sebesar 79,23%, presisi sebesar 80,18%, dan *recall* sebesar 96,39%, yang berarti bahwa model cukup baik dalam mengklasifikasikan remaja putri yang terdiagnosis anemia dan tidak anemia.

Kata Kunci: Anemia, Klasifikasi, Regresi Logistik

1. PENDAHULUAN

Anemia merupakan masalah kesehatan yang serius, terutama di negara-negara sedang berkembang, dan dapat mengakibatkan berbagai komplikasi jika tidak ditangani dengan baik. Berdasarkan data *World Health Organization* (WHO) yang menyatakan bahwa prevalensi anemia dunia mencapai 40% hingga 80%, dimana angka kejadian sebesar 53,7% terjadi pada remaja putri. Sedangkan di Indonesia, angka kejadian anemia terbilang masih tinggi. Data Riskedas (Riset Kesehatan Dasar) menyatakan bahwa kejadian anemia pada remaja putri mencapai 27,2% yang mengindiskan perempuan memiliki peluang terkena anemia yang lebih besar. Prevalensi anemia kelompok remaja sebesar 32% dari 27,2% tersebut, yang berarti bahwa 3 hingga 4 dari 10 remaja putri menderita anemia [1].

Menurut WHO, usia remaja putri adalah perempuan yang berusia 15 – 25 Tahun, dimana usia tersebut remaja putri rentan terkena anemia karena mengalami menstruasi sehingga

cenderung mengalami penurunan zat besi yang ada didalam tubuh [2].

Anemia juga dapat disebabkan oleh penyakit kanker, malaria, hepatitis, dan diare [3],[4]. Meskipun anemia terutama dipengaruhi oleh konsentrasi zat besi, *ferritin serum*, dan *Soluble Transferin Receptor* (STfR) juga dapat mempengaruhi anemia melalui hubungannya dengan reaksi peradangan. CRP, *ferritin serum*, dan STfR merupakan protein fase akut lain yang berfungsi sebagai penanda awal peradangan atau infeksi dan juga menyebabkan penurunan kadar hemoglobin dalam tubuh [5]. Berdasarkan uraian di atas dapat dilihat bahwa kejadian anemia dipengaruhi oleh faktor defisiensi zat besi, *ferritin serum*, STfR dan riwayat penyakit kronik. Faktor-faktor tersebut memiliki pengaruh yang berbeda-beda. Sehingga untuk memudahkan melihat pola hubungan antara faktor-faktor tersebut dengan kejadian anemia dapat dilakukan penerapan teknik data *mining*.

Data *mining* sering digunakan dalam suatu proses untuk memperoleh informasi dari *dataset* yang digunakan dalam pengambilan suatu keputusan. *Data mining* menyatukan teknik pembelajaran mesin, statistika, pengenalan pola visualisasi dan *database* dalam menangani permasalahan dalam pengambilan informasi dari penyimpanan *database* yang besar. *Data mining* memiliki beberapa teknik dalam pengerjaannya dan juga untuk menemukan pola ataupun informasi yang tersembunyi. Salah satu metode yang dapat diselesaikan pada masalah *data mining* adalah klasifikasi.

Klasifikasi menjadi salah satu teknik yang digunakan agar dapat memperkirakan atau mengetahui kelas pada suatu objek berdasarkan atribut yang tersedia. Klasifikasi dapat digunakan pada berbagai bidang antara lain bidang perbankan, kesehatan, perdagangan dan

juga industri. Penyelesaian kasus klasifikasi dapat diolah menggunakan metode statistika. Salah satu metode yang dapat digunakan dalam penyelesaian masalah klasifikasi yaitu analisis regresi logistik. Penggunaan regresi logistik dilakukan mengetahui ada atau tidaknya hubungan diantara dua variabel atau lebih. Variabel yang digunakan merupakan variabel dependen kategori yang bersifat binary/dikotom dan variabel independen yang bersifat kategori atau numerik. Analisis regresi logistik dapat digunakan dalam menyelesaikan masalah klasifikasi.

Berdasarkan uraian latar belakang maka penelitian ini bertujuan untuk mengetahui model klasifikasi penyakit anemia menggunakan metode regresi logistik. Penelitian ini akan memberikan wawasan baru dalam pemanfaatan *data mining* dalam kasus klasifikasi *dataset* menggunakan regresi logistik.

2. TINJAUAN PUSTAKA

Pengantar *Machine Learning*

Machine learning merupakan suatu teknik dalam bidang kecerdasan buatan (*Artificial Intelligence*) yang terfokus pada algoritma komputer untuk belajar dari data dan pengalaman untuk melakukan tugas tertentu. Metode *machine learning* dapat dibedakan menjadi *Supervised Learning* (Pembelajaran terawasi) dan *Unsupervised Learning* (Pembelajaran tidak terawasi) [6].

Supervised learning merupakan suatu teknik di dalam *machine learning* yang menggunakan data input dan output yang telah diketahui pada tahap sebelumnya. Tujuannya adalah untuk mengembangkan model yang dapat memprediksi output (label) yang tepat berdasarkan pada data input yang diberikan. Beberapa contoh metode yang termasuk dalam *supervised learning* adalah *Support Vector Machine*, *Rule-based Methods*, dan *Decision Tree based Methods*.

Klasifikasi

Klasifikasi merupakan salah satu metode dalam pembelajaran yang mempunyai fungsi dalam menentukan setiap himpunan atribut pada satu objek. Klasifikasi juga diartikan sebagai

sebuah proses dalam pengelompokan data, dimana klasifikasi menempatkan objek pada kelas yang memiliki fungsi yang sama [7].

Metode klasifikasi memiliki kemiripan dengan estimasi, namun berbeda dari sisi variabelnya yang memiliki tipe kategorial. Dalam klasifikasi, dibagi menjadi kategori target. Misalnya dalam variabel pendapatan yang dapat digolongkan menjadi tiga kelas/kategori yakni berpendapatan tinggi, menengah dan rendah.

Proses dalam klasifikasi data dilakukan dengan cara mencocokkan data baru atau biasa disebut data *testing* dengan data latih atau data *training*. Dari hasil pencocokan yang dilakukan, maka akan memunculkan sebuah keputusan dari hasil yang telah ditargetkan dari data *training*. Banyaknya jumlah dari data *training* sangat berpengaruh terhadap keputusan yang diambil dalam proses *data testing*. Apabila jumlah data *training* besar, maka keputusan yang didapatkan akan memiliki tingkat akurasi yang tinggi. Begitu pula sebaliknya, apabila jumlah data *training* yang digunakan kecil, maka tingkat akurasi yang diperoleh akan rendah.

Regresi Logistik

Analisis regresi merupakan salah satu teknik statistika yang dipergunakan untuk dapat mengetahui hubungan antara dua atau lebih variabel, salah satunya adalah analisis regresi logistik. Analisis ini merupakan salah satu pendekatan matematis yang bisa digunakan dalam menganalisis hubungan satu atau beberapa variabel independen dengan variabel dependen kategori yang bersifat binary/dikotom [8].

Jika variabel dependen terdiri dari dua jenis kategorik, maka disebut sebagai regresi logistik biner, bernilai 1 apabila kejadian sukses dan bernilai 0 apabila kejadian gagal. Adapun model dari peluang antara variabel independen $X_{i1}, X_{i2}, \dots, X_{ip}$ dengan variabel dependen (Y), Dimana peluang kategorinya (π) yaitu sebagai berikut [9].

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})} \quad (2.1)$$

dimana,

$\pi(x_i)$: Peluang kategori variabel dependen

X_{pi} : Variabel independen ke – p , responden ke- i

p : Banyaknya variabel independen

β_0 : Intercept

$\beta_1, \beta_2, \dots, \beta_p$: Koefisien regresi untuk setiap variabel independen

Agar dapat mempermudah interpretasi serta pendugaan parameternya, maka peluang pada persamaan (2.1) di atas ditransformasikan sehingga diperoleh fungsi logit sebagai berikut.

$$g(x_i) = \text{logit } \pi(x_i) = \ln \frac{\pi(x_i)}{1-\pi(x_i)} \quad (2.2)$$

Jika,

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}$$

$$g(x_i) = \beta_0 + \sum_{j=1}^p \beta_j X_{ji} \quad (2.3)$$

Penggunaan regresi logistik meliputi beberapa langkah-langkah hingga diperoleh hasil klasifikasi yang diinginkan. Prosedur analisis klasifikasi menggunakan regresi logistik dijelaskan sebagai berikut [10].

1. Pendugaan Parameter

Metode yang dipergunakan dalam menduga nilai parameter pada Regresi logistik yaitu metode *Maximum Likelihood Estimation* (MLE). Dalam metode ini, β dapat digunakan dengan memaksimalkan fungsi *likelihood*. Adapun fungsi *likelihood* pada model regresi logistik yaitu sebagai berikut [11].

$$L(\beta) = \prod_{i=1}^n f(y_i|\beta) \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (2.4)$$

2. Asumsi Multikolinieritas

Multikolinieritas menunjukkan terjadinya korelasi dari beberapa atau semua variabel independen. Multikolinieritas yang tinggi dapat mengindikasikan ketergantungan antar dua atau lebih dari variabel independen. Proses pengujian dilakukan dengan hipotesis sebagai berikut.

H_0 : Tidak terdapat multikolinieritas

H_1 : Terdapat multikolinieritas

Multikolinieritas dapat diketahui dengan melihat *Variance Inflation Factor* (VIF), dan dapat dihitung dengan rumus sebagai berikut [12].

$$VIF_j = \frac{1}{1-R_j^2} \quad (2.5)$$

dimana,

R_j^2 : Koefisien determinasi

Nilai R_j^2 berbanding lurus dengan nilai VIF.

Jika nilai R_j^2 semakin besar, maka nilai VIF akan semakin besar pula. Apabila nilai VIF bernilai lebih dari 10, maka diterima H_1 atau dapat dikatakan bahwa data yang digunakan terdapat multikolinieritas.

3. Uji Simultan

Uji simultan digunakan untuk mengetahui adanya pengaruh variabel independen terhadap variabel dependen secara simultan atau bersama-sama. Adapun hipotesis yang digunakan pada uji signifikansi parameter secara simultan adalah sebagai berikut.

H_0 : $\beta_1 = \beta_2 = \dots = \beta_j = 0; j = 1, 2, \dots, p$

H_1 : Minimal terdapat satu $\beta_j \neq 0$.

Adapun statistik uji yang digunakan adalah sebagai berikut.

$$G^2 = -2 \ln \left(\frac{L_0}{L_1} \right) \quad (2.6)$$

Dimana,

G^2 : *likelihood ratio test*

L_0 : nilai maximum *likelihood* untuk model tanpa variabel independen

L_1 : nilai maximum *likelihood* untuk model dengan variabel independen

Uji G^2 mengikuti sebaran χ^2 pada derajat bebas p , dimana p yaitu banyak variabel independen. H_0 ditolak apabila nilai $G \geq \chi^2_{(p)}$ [13].

4. Uji Parsial

Uji parsial digunakan agar dapat mengetahui pengaruh dari masing-masing variabel independen terhadap variabel dependennya. Hipotesis yang digunakan adalah

H_0 : $\beta_j = 0 ; j = 1, 2, \dots, p$

H_1 : $\beta_j \neq 0$

Statistik uji yang digunakan adalah.

$$W = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (2.7)$$

Dimana,

$\hat{\beta}_j$: Pendugadari β_j

$se(\beta_j)$: Standar *error* dari penduga β_j

Statistik uji mengikuti sebaran normal. H_0 akan ditolak jika nilai $W > Z_{\alpha/2}$ (Setiap variabel independen tidak berpengaruh terhadap variabel dependen) [14].

5. Uji Kesesuaian Model

Uji kesesuaian model dipergunakan mengevaluasi model yang telah diperoleh telah sesuai dengan data, nilai yang dihasilkan sama atau mendekati dengan yang diinginkan dalam model. Statistik uji yang digunakan dalam uji kesesuaian model adalah *Uji Pearson*.

Adapun hipotesis yang digunakan adalah sebagai berikut.

H_0 : Model yang diperoleh telah sesuai

H_1 : Model yang diperoleh tidak sesuai

Rumus Uji *Pearson* [15].

$$\chi^2_{Pearson} = \sum_{i=1}^n r(Y_i, \pi_i)^2 \sim \chi^2_{(n-p-1)}$$

$$= \sum_{i=1}^n \left(\frac{Y_i - n_i \pi_i}{\sqrt{n_i \pi_i (1 - \pi_i)}} \right)^2 \sim \chi^2_{(n-p-1)}$$

(2.8)

dimana,

Y_i : Variabel dependen pengamatan ke- i

n_i : Total pengamatan ke- i

π_i : Peluang variabel dependen pada pengamatan ke- i

Model akan disebut sesuai jika uji $\chi^2_{Pearson}$ tidak lebih besar dari nilai $\chi^2_{(n-p-1)}$.

6. Klasifikasi Pada Regresi Logistik

Klasifikasi dapat diselesaikan dengan menggunakan regresi logistik. Pendekatan dengan regresik logistik dalam memprediksi kelas klasifikasi dilakukan dengan menghitung probabilitasnya. Klasifikasi yang berasal dari variabel dependennya didapatkan dengan cara menentukan nilai titik potong dengan nilai sebesar 0,5. Klasifikasi yang dilakukan berdasarkan analisis regresi logistik memakai model peluang pada ketentuan yaitu.

$$\pi(x_i) = \begin{cases} 0, & \pi(x_i) < 0,5 \\ 1, & \pi(x_i) \geq 0,5 \end{cases}$$

Jika hasil peluang yang diperoleh dari model bernilai lebih kecil 0,5 maka prediksinya bernilai kategori 0. Begitu pula sebaliknya.

Jika hasil peluang yang diperoleh dari model bernilai lebih besar atau sama dengan 0,5 maka prediksinya bernilai kategori 1.

Tingkat Ketepatan Klasifikasi

Untuk mengukur hasil prediksi klasifikasi, terdapat beberapa ukuran yang digunakan yaitu nilai *Apparent Error Rate* (APER), Tipe kesalahan dan tingkat akurasi klasifikasi.

1. Nilai *Apparent Error Rate*

Nilai APER merupakan proporsi dari banyak individu yang tidak tepat dalam klasifikasi. Metode yang memiliki nilai APER yang paling kecil dapat dikatakan sebagai metode yang tingkat ketepatan klasifikasi yang baik. Adapun perhitungan nilai APER dapat dilihat pada Tabel 1 di bawah ini.

Tabel 2.1 Perhitungan Nilai APER

Prediksi	Observasi		Total
	0	1	
0	n_{11}	n_{12}	$n_{11} + n_{12}$
1	n_{21}	n_{22}	$n_{21} + n_{22}$
Total	$n_{11} + n_{21}$	$n_{12} + n_{22}$	$n_{11} + n_{21} + n_{12} + n_{22}$

Dimana n_{11} dan n_{22} adalah banyak individu yang benar dalam klasifikasi. Sedangkan n_{12} dan n_{21} adalah banyak individu yang kurang tepat dalam klasifikasi. Adapun rumus dalam menghitung nilai APER adalah sebagai berikut [16].

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{21} + n_{12} + n_{22}} \times 100\% \quad (2.9)$$

2. Tipe Kesalahan Klasifikasi

Pada konteks statistika dan teori pengujian hipotesis, kesalahan tipe I dan kesalahan tipe II merupakan dua jenis kesalahan yang mungkin terjadi ketika menguji hipotesis statistik.

a. Kesalahan Tipe I

Kesalahan tipe I merupakan peluang menolak hipotesis nol tapi pada kenyataannya hipotesis nol itu benar. Kesalahan ini juga disebut sebagai *false positive* (n_{21}). Adapun rumus menghitung tingkat kesalahan tipe I sebagai berikut.

$$\text{Tingkat Kesalahan Tipe I} = \frac{n_{21}}{n_{11} + n_{21} + n_{12} + n_{22}} \quad (2.10)$$

b. Kesalahan Tipe II

Kesalahan tipe II merupakan peluang untuk tidak menolak hipotesis nol tapi pada kenyataannya hipotesis tersebut pada kenyataannya salah. Kesalahan ini juga

disebut dengan *false negative* (n_{12}). Adapun rumus menghitung tingkat kesalahan tipe II sebagai berikut.

$$\text{Tingkat Kesalahan Tipe II} = \frac{n_{12}}{n_{11} + n_{21} + n_{12} + n_{22}} \quad (2.11)$$

Konsep kesalahan tipe I dan II dalam konteks pengujian hipotesis dapat dilihat pada Tabel 2 sebagai berikut.

Tabel 2.2 Tipe Kesalahan Hipotesis

Keputusan	Keadaan Sebenarnya	
	H ₀ Salah	H ₀ Benar
Menolak H ₀	n_{11}	Kesalahan Tipe I (n_{12})
Gagal Menolak H ₀	Kesalahan Tipe II (n_{21})	n_{22}

Keterangan:

- n_{11} : hipotesis nol salah dan keputusannya menolak hipotesis nol
- n_{12} : hipotesis nol benar dan keputusannya menolak hipotesis nol
- n_{21} : hipotesis nol salah dan keputusannya gagal menolak hipotesis nol
- n_{22} : hipotesis nol benar dan keputusannya gagal menolak hipotesis nol

3. Tingkat Akurasi Klasifikasi

Confusion matrix dapat digunakan untuk menentukan tingkat akurasi hasil klasifikasi. *Confusion Matrix* adalah teknik yang digunakan untuk melihat gambaran detail dari jumlah data yang di klasifikasi dimana fungsinya untuk mengevaluasi model klasifikasi sehingga bisa dilihat perkiraan objek yang terklasifikasi benar maupun yang terklasifikasi salah. Adapun perhitungan *confusion matrix* dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 2.3 Perhitungan *Confusion Matrix*

Prediksi	Aktual	
	0	1
0	n_{11}	n_{12}
1	n_{21}	n_{22}

Keterangan:

- n_{11} : Jumlah observasi yang tepat klasifikasi sebagai kategori 0

n_{12} : Jumlah observasi yang tidak tepat klasifikasi sebagai kategori 1

n_{21} : Jumlah observasi yang tidak tepat klasifikasi sebagai kategori 0

n_{22} : Jumlah observasi yang tepat klasifikasi sebagai kategori 1

Suatu model dikatakan baik harus mempunyai nilai n_{11} dan n_{22} yang tinggi dan juga kesalahan klasifikasinya harus lebih kecil. Kebaikan model klasifikasi bisa dilihat juga pada akurasi, presisi, dan *recall* yang memiliki nilai tinggi. Akurasi adalah ukuran yang digunakan untuk mengetahui banyaknya prediksi yang tepat klasifikasi berdasarkan pada model yang terbentuk, presisi digunakan untuk mengukur sejauh mana model dapat mengidentifikasi positif dengan benar dari semua contoh yang diprediksi positif oleh model, dan *recall* digunakan dalam mengukur sejauh mana model dapat mengidentifikasi positif dengan benar dari semua contoh positif dalam dataset. Adapun persamaannya yaitu sebagai berikut [17].

$$\text{Accuracy} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (2.12)$$

$$\text{Precision} = \frac{n_{11}}{n_{11} + n_{12}} \quad (2.13)$$

$$\text{Recall} = \frac{n_{11}}{n_{11} + n_{21}} \quad (2.14)$$

3. METODOLOGI

Penelitian dilaksanakan pada bulan Februari 2023 hingga Oktober 2023. Jenis penelitian ini adalah penelitian terapan yaitu menerapkan teknik data *mining* untuk mengklasifikasi faktor-faktor penyebab anemia pada remaja wanita.

Prosedur Analisis

Adapun langkah-langkah analisis yang diterapkan oleh peneliti adalah sebagai berikut:

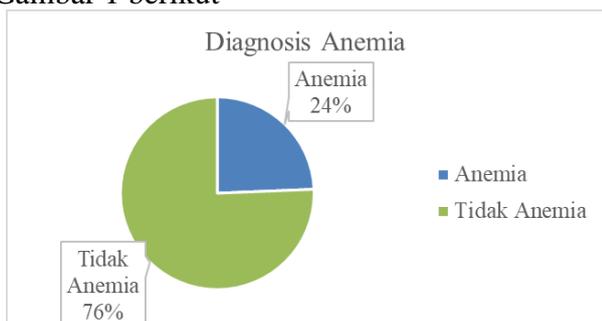
1. Mempersiapkan data yang akan diklasifikasi dengan membagi data menjadi dua bagian yakni data *training* dan data *testing*. Proporsi yang digunakan yaitu data *training* sebanyak 85% dan data *testing* sebanyak 15%.
2. Pemodelan regresi logistic yang meliputi beberapa tahapan yaitu.
 - a. Menduga parameter
 - b. Memeriksa asumsi multikolinearitas
 - c. Menguji signifikansi parameter

- d. Membentuk model logistik menggunakan variabel yang signifikan
 - e. Melakukan klasifikasi berdasarkan model yang dihasilkan
3. Melakukan evaluasi model klasifikasi dengan menghitung akurasi, tipe *error*, presisi dan *recall*.

4. HASIL DAN PEMBAHASAN

Pembagian Data *Training* dan *Testing*

Data yang digunakan pada penelitian ini yaitu data anemia remaja putri dengan rentang usia 12-15 Tahun. Setelah melakukan data *selection*, *preprocessing*, dan transformasi dataset, data anemia remaja putri di Indonesia diperoleh dataset sebesar 2.444 data dimana hasil diagnosis tidak anemia sebanyak 1.849 (76%) dan diagnosis anemia sebanyak 595 (24%). Adapun presentasi data anemia dapat dilihat pada Gambar 1 berikut



Gambar 4.1 Persentase Jumlah Diagnosa Penyakit Anemia

Selanjutnya proses pembentukan model regresi logistik dilakukan dengan menggunakan 85% data *training* atau sebanyak 2.078 data, model yang diperoleh akan digunakan dalam proses klasifikasi pada data *testing*. Adapun langkah-langkah yang dilakukan dalam membentuk model regresi logistik diantaranya melakukan pendugaan parameter, pendeteksian multikolinearitas, pengujian signifikansi parameter, uji kesesuaian model dan menghitung ketepatan klasifikasi.

Membentuk Model Regresi Logistik

Selum melakukan Klasifikasi, langkah awal yang dilakukan yaitu membentuk model regresi logistic. Adapun langkah-langkah adalah sebagai berikut.

1. Pendugaan Parameter

Hasil pendugaan parameter dengan menggunakan *Maximum Likelihood Estimation* untuk seluruh variabel independen yaitu sebagai berikut.

Tabel 4.1 Penduga Parameter

Koefisien	Estimasi
<i>Intercept</i>	-1,5468
X_1 (<i>Ferritin Serum</i>)	0,8812
X_2 (<i>Transferrin Receptor</i>)	1,1255
X_3 (Riwayat Kanker)	14,4422
X_4 (Diagnosa Diare)	-0,3293
X_5 (Diagnosa Hepatitis)	-0,6509
X_6 (Diagnosa Malaria)	0,2055

Berdasarkan Tabel 4 di atas, maka model yang terbentuk untuk status anemia adalah sebagai berikut.

$$g(x) = -1,5468 + 0,8812X_1 + 1,1255X_2 + 14,4422X_3 + -0,3293X_4 - 0,6509X_5 + 0,2055X_6$$

2. Pendeteksian Multikolinearitas

Pendeteksian multikolinearitas pada analisis regresi logistik dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF) untuk setiap variabel independen. Adapun hasil nilai VIF untuk setiap variabel dapat dilihat pada Tabel 4.2 berikut.

Tabel 4.2 Nilai VIF Variabel Independen

Variabel	Nilai VIF
X_1	1,0945
X_2	1,0962
X_3	1,0000
X_4	1,0152
X_5	1,0167
X_6	1,0294

Berdasarkan Tabel 4.2 di atas, dapat disimpulkan bahwa nilai VIF pada kasus klasifikasi pada penyakit anemia untuk semua variabel independen memperlihatkan hasil yang bernilai lebih kecil dari 10 sehingga tidak terdapat multikolinieritas.

2. Pengujian Signifikansi Parameter

Uji signifikansi parameter terdiri dari dua jenis uji, yakni uji simultan dan uji parsial. Untuk uji simultan, dilakukan agar dapat mengetahui pengaruh antara variabel independen dengan variabel dependen secara bersama-sama.

Hasil uji signifikansi parameter secara simultan dengan melihat nilai Statistik uji G.

Dalam penelitian ini, hasil nilai statistik Uji G sebesar 12,5915 yang lebih besar dari nilai titik kritisnya yaitu sebesar 1,7059 maka diputuskan tolak H_0 dan diterima H_1 sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel independen yang memiliki pengaruh terhadap variabel dependen.

Selanjutnya, dilakukan uji parsial agar dapat diketahui pengaruh masing-masing variabel independen terhadap variabel dependen yang dilakukan. Uji parsial akan dilakukan menggunakan.

Uji signifikansi secara parsial dilakukan dengan membandingkan nilai p -value masing-masing variabel dengan taraf signifikan sebesar 0,05. Kesimpulan yang diambil adalah tolak H_0 jika nilai p -value bernilai kurang dari 0,05 dan begitu pula sebaliknya. Hasil nilai uji parsial dari masing-masing variabel dapat dilihat pada tabel 4.3 berikut.

Tabel 4.3 Nilai Uji Signifikansi Parsial

Koefisien	P-Value	Kesimpulan
Intercept	$2e^{-16}$	Signifikan
X_1	$1,90e^{-13}$	Signifikan
X_2	$2,61e^{-14}$	Signifikan
X_3	0,965	Tidak Signifikan
X_4	0,116	Tidak Signifikan
X_5	0,215	Tidak Signifikan
X_6	0,304	Tidak Signifikan

Tabel 4.3 di atas memperlihatkan bahwa variabel *Ferritin Serum* dan *Transferin Receptor* yang berpengaruh secara signifikan terhadap status penyakit anemia, sedangkan variabel lainnya tidak berpengaruh secara signifikan.

3. Uji Kesesuaian Model

Nilai uji *Pearson* yang dihasilkan adalah 3,0721 dan nilai p -value-nya adalah 0,3806 lebih besar dari nilai α sehingga dapat H_0 diterima, yang berarti model regresi logistik yang diperoleh layak untuk digunakan.

Evaluasi Model

Model yang diperoleh pada data *testing* selanjutnya diuji untuk mengetahui seberapa baik model tersebut dalam menghitung ketepatan klasifikasi pada data *testing*. Kebaikan model dapat dilihat dari nilai akurasi, tipe *error*, presisi,

dan *recall*. Adapun hasil klasifikasi pada data *testing* dapat dilihat pada Tabel 4.4 berikut.

Tabel 4.4 Ketetapan Klasifikasi untuk Data *Testing*

Prediksi	Kategori		Total
	0	1	
0	267	66	333
1	10	23	33
Total	277	89	366

Tabel 4.4 di atas memperlihatkan bahwa pada klasifikasi menggunakan regresi logistik terdapat 267 pengamatan yang tepat dalam klasifikasi untuk kategori 0, yakni orang yang tidak memiliki status penyakit anemia dan terdapat 66 pengamatan yang salah klasifikasi dimana untuk kategori 0 menghasilkan hasil klasifikasi kategori 1. Selanjutnya terdapat 23 pengamatan yang tepat dalam klasifikasi untuk kategori 1, yakni orang yang memiliki status penyakit anemia dan terdapat 10 pengamatan yang salah klasifikasi dimana untuk kategori 1 menghasilkan klasifikasi kategori 0.

Selanjutnya dilakukan evaluasi hasil klasifikasi dengan menghitung tipe kesalahan, *accuracy*, *precision*, dan *recall* pada regresi logistik sebagai berikut.

$$Accuracy = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% = \frac{267 + 23}{267 + 10 + 66 + 23} \times 100\% = 79,23\%$$

$$Precision = \frac{n_{11}}{n_{11} + n_{12}} = \frac{267}{267 + 66} = 0,8018 = 80,18\%$$

$$Recall = \frac{n_{11}}{n_{11} + n_{21}} = \frac{267}{267 + 10} = 0,9639 = 96,39\%$$

Berdasarkan data hasil pengujian *confusion matrix* untuk model klasifikasi menggunakan regresi logistik, diketahui bahwa pengujian *confusion matrix* untuk model klasifikasi menggunakan regresi logistik yaitu hasil akurasi klasifikasi regresi logistik sebesar 79,23% yang berarti model yang dihasilkan mampu memprediksi data dengan tepat sebanyak 290 dari data *testing* sebanyak 366. Nilai presisi didapatkan sebesar 80,18% artinya pada model yang dihasilkan memberikan kemampuan yang baik dalam mengukur keakuratan prediksi positif dengan tingkat kesalahan untuk kelas positif

sebesar 19,82% dan nilai *recall* sebesar 96,39% yang berarti bahwa model yang dihasilkan memiliki kemampuan yang baik untuk mengidentifikasi dan menemukan semua kelas positif yang sebenarnya.

5. KESIMPULAN

Berdasarkan hasil penelitian maka disimpulkan bahwa dengan menggunakan model regresi logistic diperoleh model terbaik yang menghubungkan antara variabel independen dengan status anemia dimana variabel independen yang berpengaruh secara signifikan ialah variabel *ferritin serum* dan *transperin receptor*, dengan model sebagai berikut

$$g(x) = -1,5468 + 0,8812X_1 + 1,1255X_2$$

Berdasarkan persamaan regresi tersebut diperoleh kinerja evaluasi model, dimana terdapat 23 pengamatan yang benar diprediksi anemia dan 267 pengamatan yang benar diprediksi tidak anemia dengan akurasi sebesar 79,23%, *sensitivity* sebesar 96,39%, dan *precision* sebesar 80,18%. Yang berarti bahwa model memiliki kinerja yang cukup baik dalam mengklasifikasikan data.

6. DAFTAR PUSTAKA

- [1] Efriwati. (2018). *Hasil Utama RISKESDAS*. https://kesmas.kemkes.go.id/assets/upload/dir_519d41d8cd98f00/files/Hasil-riskesdas-2018_1274
- [2] WHO. (2017). *Worldwide Prevalance of Anemia*.
- [3] Azmi, S.K., (2005). Pengaruh Anemia pada Kanker Terhadap Kualitas Hidup dan Hasil Pengobatan. Universitas Sumatera Utara, Medan.
- [4] Monica, M.M., (2014). Hubungan Antara Malaria Klinis dengan Anemia pada Penderita yang Berkunjung di Puskesmas Wori Kabupaten Minahasa Utara. *Jurnal Kedokteran Komunitas dan Tropik*. Vol 11, No.1.
- [5] Aidi, M. N., Ernawati, F., Efriawati, Nurjannah, N. Rachmawati, R. Julianti, E. D., Sundari, D., Retiaty, F. Fitrianto, A. Nurfadilah, K. Arifin, A. Y. (2022). *Spatial distribution and identifying biochemical factors affecting haemoglobin levels among women of reproductive age for each province in Indonesia: A geospatial analysis*. *Geospatial Health* 2022; Vol. 17:1118
- [6] Brownlee, J. (2016). Master Machine Learning Algorithms Discover. In *How They Work and Implement Them Fron Scratch*.
- [7] Putry, N. M. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus. *EVOLUSI: Jurnal Sains Dan Manajemen*, 10(1). <https://doi.org/10.31294/evolusi.v10i1.12514>
- [8] Alwi, W., Ermawati, & Husain, S. (2018). Analisis Regresi Logistik Biner untuk Memprediksi Kepuasan Pengunjung Pada Rumah Sakit Umum Daerah Majene. *Journal MSA*, 20-26.
- [9] Hosmer, D. ., & Lemeshow, S. (2000). *Applied Logistic Regression* (Second Edi). Willey-Interscience : Canada.
- [10] Harlan, J. (2018). *Analisis Regresi Logistik*. Gunadharma : Depok.
- [11] Novianti, F. A., & Purnami, S. W. (2012). Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal SAINS Dan Seni ITS*, 1(1), D147–D152.
- [12] Gujarati, D. N. (2004). *Basic Econometric* (Fourth Edi). Gary Burke : North America.
- [13] Moni, M. P., Ginting, K. B., & Aryanto. (2016). Analisis Regresi Logistik Terhadap Faktor-Faktor Yang Mempengaruhi Kejadian Malaria Pada Balita Di Kecamatan Lamboya Kabupaten Sumba Barat. *Jurnal MIPA FST UNDANA*, Vol 20, No. 2,3-4.
- [14] Zulkifli, M.J.M. (2014). *Pendekatan Regresi Logistik Multinomial Pada Klasifikasi Pemilihan Jurusan Siswa SMA 5 Negeri Malang*. Jurusan Matematika Universitas Brawijaya: Malang.
- [15] Antika, D. (2017). *Klasifikasi Menggunakan Metode Regresi Logistik dan Support Vector Machine*. Universitas Brawijaya : Malang.
- [16] Ndangi, W. R. A, Resmawan., & Djakaria, I. (2019). Perbandingan Analisis Diskriminan dan Regresi Multinomial.

Jambura Journal Of Mathematics, Vol. 1,
No. 2, 54-63.

- [17] Larner, A. (2021). *Contingency, Confusion and the Metrics of Binary classification*. Liverpool: Walton Centre for Neurology and Neurosurgery.