

Prediksi Ketepatan Klasifikasi Status Predikat Lulusan Program Sarjana FMIPA Universitas Mulawarman Menggunakan Regresi Logistik Biner dan *Neural Networks*

Prediction of Predicate Status for Graduates of Undergraduate Programs, Faculty of Mathematics and Natural Sciences, Mulawarman University Using Binary Logistic Regression and Neural Networks

Lisa Dwi Nurul Khasanah¹, M. Fathurahman^{2a)}, and Memi Nor Hayati³

^{1,2,3}Program Studi Statistika, Jurusan Matematika, FMIPA Universitas Mulawarman, Indonesia

^{a)}Corresponding author: fathur@fmipa.unmul.ac.id

ABSTRACT

Classification is a learning technique for identifying categorical groups from a data set whose group member categories are known. Several methods that can be used in classification include binary logistic regression and neural networks. This research aims to compare the prediction results for the accuracy of the classification of predicate status for graduates of the FMIPA Mulawarman University undergraduate program in 2021. In the binary logistic regression method, the model parameters are estimated using the maximum likelihood estimation and Fisher scoring iteration methods. The neural networks used the backpropagation algorithm. The results of the research show that the classification accuracy using the confusion matrix obtained with binary logistic regression and neural networks is the same, namely 87.5%.

Keywords: binary logistic regression, classification, confusion matrix, neural networks

1. Pendahuluan

Metode regresi merupakan teknik analisis data yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Ada berbagai jenis metode regresi yang dikenal saat ini, salah satunya jika variabel respon berupa data kategorik, maka dapat dimodelkan dengan menggunakan regresi logistik (Hosmer et al, 2013). Jika dilihat dari variabel responnya, regresi logistik terdiri atas regresi logistik biner, regresi logistik multinomial, dan regresi logistik ordinal (Fatonah dkk, 2019). Dalam penelitian ini, digunakan regresi logistik biner, karena variabel responnya adalah status predikat lulusan program sarjana yang dibedakan menjadi dua kategori tingkatan.

Penelitian sebelumnya yang berkaitan dengan faktor-faktor yang memengaruhi predikat lulusan sarjana telah dilakukan oleh Imaslihkah dkk (2013) dengan faktor-faktor yang memengaruhi predikat lulusan sarjana di Institut Teknologi Sepuluh Nopember Surabaya diantaranya fakultas, jenis kelamin, asal daerah, jalur penerimaan, status SMA, pekerjaan ayah, pekerjaan ibu, dan pendapatan. Nurrohmat dan Nugroho (2015) dalam penelitiannya memprediksi masa studi dan predikat lulusan sarjana di Universitas Muhammadiyah Surakarta dengan faktor-faktor yang memengaruhi yaitu jurusan SMA, jenis kelamin, asal daerah, asal sekolah dan asisten praktikum.

Penelitian mengenai regresi logistik juga dapat digabungkan bersama metode lainnya seperti yang dilakukan oleh Ramli dkk (2013), yaitu membandingkan antara regresi logistik dan *neural networks* pada pengklasifikasian pemilihan jurusan Bahasa dan IPS pada Sekolah Menengah Atas (SMA) Negeri 2 Samarinda. Metode *neural networks* memiliki ketepatan klasifikasi yang lebih baik dibanding regresi logistik (Ramli dkk, 2013). Hasanah (2019) membandingkan klasifikasi menggunakan *neural networks* dan regresi logistik dengan hasil yang didapat menunjukkan bahwa ketepatan klasifikasi dengan menggunakan *confusion matrix* pada *neural networks* lebih baik dibanding regresi logistik.

Neural networks merupakan suatu sistem yang konsepnya meniru sistem yang ada pada tubuh manusia. Menurut Sutojo dkk (2011), *neural networks* memiliki kelebihan dan kekurangan. Kelebihan dari *neural networks* adalah mampu melakukan pembelajaran berdasarkan pengalaman awal yang dimiliki, mampu membuat organisasi secara tersendiri berdasarkan informasi yang diterima selama melakukan pembelajaran, dan dapat melakukan proses perhitungan secara paralel yang dapat mempersingkat proses yang dilakukan.

Pada berbagai permasalahan, dibutuhkan analisis yang melibatkan proses klasifikasi. Klasifikasi bertujuan mengelompokkan objek pengamatan berdasarkan serangkaian variabel-variabel yang disebut sebagai variabel prediktor. Proses klasifikasi diperlukan data yang dibagi menjadi data *training* dan data *testing*. Pada penelitian yang pernah dilakukan oleh Hasanah (2019), meneliti klasifikasi dengan pembagian proporsi 70% data *training* dan 30% data *testing*. Pembagian proporsi 70% data *training* dan 30% data *testing* yang dilakukan agar algoritma dalam pengklasifikasian dapat lebih terlatih dan hasil pelatihan yang diujikan mendapatkan hasil yang akurat.

Berdasarkan latar belakang tersebut, penulis tertarik untuk melakukan penelitian dengan judul “Prediksi Ketepatan Klasifikasi Status Predikat Lulusan Program Sarjana FMIPA Universitas Mulawarman Menggunakan Regresi Logistik Biner dan *Neural Networks*”. Tujuan yang ingin dicapai dari penelitian ini adalah untuk mendapatkan hasil klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menggunakan regresi logistik biner, untuk mendapatkan hasil klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menggunakan *neural networks*, dan untuk mendapatkan hasil perbandingan nilai akurasi antara regresi logistik biner dan *neural networks* pada klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman.

2. Tinjauan Pustaka

2.1 Regresi Logistik Biner

Regresi logistik biner digunakan untuk memodelkan suatu kejadian dengan variabel respon dengan tipe kategori dua pilihan yaitu sukses atau gagal yang dinotasikan dengan $Y = 1$ (Sukses) dan $Y = 0$ (Gagal). Distribusi yang digunakan regresi logistik biner adalah distribusi *Bernoulli* (Agresti, 2019).

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

dengan y_i adalah peluang peubah acak ke- i .

Menurut Varamita (2017), model regresi logistik diasumsikan bahwa variabel biner harus saling bebas, sehingga variabel biner memiliki sebaran binom. Model regresi logistik dengan sebaran binom dapat dituliskan sebagai berikut:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$

atau

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

dengan

p : banyak variabel prediktor

$\pi(x)$: probabilitas sukses dengan nilai probabilitas $0 \leq \pi(x) \leq 1$.

Nilai probabilitas $\pi(x)$ berkisar antara nol dan satu, linier sederhana representasi untuk π atas semua nilai x yang mungkin tidak memadai, karena nilainya linier dalam kisaran $(-\infty, +\infty)$. Selanjutnya, transformasi harus dilakukan untuk memungkinkan nilai x memiliki nilai yang sesuai dalam kisaran $[0,1]$. Mempertimbangkan logistik transformasi disebut juga *logit* (Menezes et al, 2017). Persamaan yang diperoleh dari hasil transformasi dapat dituliskan sebagai berikut:

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

Metode yang dapat digunakan untuk mendapatkan penaksir parameter model regresi logistik biner adalah metode *Maximum Likelihood Estimation* (MLE). Menurut Hosmer et al (2013), fungsi *likelihood* berdistribusi *Bernoulli* untuk n sampel independen. Parameter β diestimasi dengan cara memaksimalkan fungsi *likelihood* yang merupakan penyelesaian dari turunan pertama dari fungsi logaritma natural *likelihood* (Aji, 2014). Fungsi *likelihood* dengan menggunakan distribusi *Bernoulli* pada Persamaan (1) adalah

$$\begin{aligned} l(\beta) &= \prod_{i=1}^n P(Y = y_i) \\ &= \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \\ &= \left\{ \prod_{i=1}^n \left(1 + \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right)^{-1} \right\} \exp \left\{ \sum_{i=1}^n \sum_{j=1}^p \beta_j x_{ij} y_i \right\} \end{aligned} \tag{4}$$

Untuk memudahkan perhitungan, maka fungsi *likelihood* dimaksimumkan dalam bentuk $\ln l(\beta)$

$$L(\beta) = \ln l(\beta)$$

$$= \ln \left\{ \prod_{i=1}^n \left(1 + \exp \left(\sum_{j=0}^p \beta_j x_{ij} \right) \right)^{-1} \right\} \exp \left\{ \sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} y_i \right\}$$

$$= \left\{ \sum_{i=1}^n \sum_{j=0}^p \beta_j x_{ij} y_i \right\} - \sum_{i=1}^n \ln \left\{ 1 + \exp \left(\sum_{j=0}^p \beta_j x_{ij} \right) \right\} \tag{5}$$

Selanjutnya, menurunkan $L(\beta)$ terhadap β_p dan hasilnya disamadengankan nol

$$\frac{\partial L(\beta)}{\partial \beta_p} = \frac{\partial \left\{ \sum_{i=1}^n \sum_{j=1}^p \beta_j x_{ij} y_i - \sum_{i=1}^n \ln \left\{ 1 + \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right\} \right\}}{\partial \beta_p}$$

$$= \sum_{i=1}^n x_{ip} y_i - \sum_{i=1}^n x_{ip} \pi(x_i) = 0 \tag{6}$$

Menurut Hosmer et al (2013), estimasi varian dan kovarian diperoleh dari turunan kedua fungsi *ln likelihood*.

$$\frac{\partial^2 L(\beta)}{\partial \beta_p^2} = \frac{\partial^2 \left\{ \sum_{i=1}^n \sum_{j=1}^p \beta_j x_{ij} y_i \right\} - \sum_{i=1}^n \ln \left\{ 1 + \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right\}}{\partial \beta_p^2}$$

$$= - \sum_{i=1}^n x_{ip}^2 \pi(x_i) (1 - \pi(x_i)) \tag{7}$$

Hasil turunan fungsi $L(\beta)$ terhadap parameter yang ditaksir berbentuk tidak eksplisit. Pendekatan numerik yang dapat digunakan untuk mendapatkan penaksir parameter yaitu dengan metode iterasi *Fisher Scoring*. Metode ini menggunakan turunan parsial kedua fungsi $L(\beta)$ terhadap parameter yang ditaksir seperti pada Persamaan (7). Persamaan yang digunakan untuk mendapatkan penaksir parameter dengan metode iterasi *Fisher Scoring* adalah:

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \mathbf{I}(\hat{\beta}_t)^{-1} \mathbf{D}(\hat{\beta}_t) \tag{8}$$

dimana,

$\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1 \ \dots \ \hat{\beta}_p]^T$ adalah parameter model yang ditaksir, $\mathbf{I}(\hat{\beta})^{-1}$ adalah matriks informasi *Fisher*, $\mathbf{D}(\hat{\beta})$ adalah vektor gradien.

Proses iterasi *Fisher Scoring* untuk parameter β akan berhenti ketika nilai taksiran yang didapat konvergen. Kondisi konvergen yaitu $\|\beta_{t+1} - \beta_t\| < \varepsilon$, dimana ε adalah bilangan yang sangat kecil. Sehingga didapatkan hasil penaksiran β_{t+1} pada saat iterasi terakhir (Fathurahman, 2017).

Pengujian parameter model regresi logistik biner meliputi uji serentak dan uji parsial. Uji serentak digunakan untuk mengetahui minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel prediktor atau untuk mengetahui adanya pengaruh variabel prediktor terhadap variabel respon secara bersama-sama (serentak). Sedangkan, uji parsial digunakan untuk mengetahui adanya pengaruh yang signifikan dari masing-masing variabel prediktor terhadap variabel respon secara individu (parsial). Hipotesis yang digunakan untuk uji serentak diformulasikan sebagai berikut:

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

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, p. \tag{9}$$

Menurut Suaniantara dkk (2019), statistik uji untuk menguji hipotesis pada Persamaan (9) yang diperoleh menggunakan fungsi *likelihood* dapat diformulasikan sebagai berikut:

$$G = -2 \ln \left[\frac{\text{likelihood tanpa variabel prediktor}}{\text{likelihood dengan variabel prediktor}} \right]$$

$$G = -2 \ln \left[\frac{\binom{n_1}{n}^{n_1} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \tag{10}$$

Atau

$$G = 2 \left\{ \sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\}$$

Keterangan :

n_0 : Banyaknya pengamatan yang bernilai $Y = 0$

n_1 : Banyaknya pengamatan yang bernilai $Y = 1$

n : Total pengamatan

Statistik uji pada Persamaan (10) berdistribusi *chi-square*. Daerah penolakan H_0 (daerah kritis) dari pengujian hipotesis pada Persamaan (9) untuk tingkat signifikansi α adalah tolak H_0 jika nilai $G > \chi^2_{(\alpha,p)}$ atau tolak H_0 bila *p-value* kurang dari α . Nilai $\chi^2_{(\alpha,p)}$ dapat diperoleh dari tabel distribusi *chi-square*.

Pengujian hipotesis parameter selanjutnya adalah uji parsial. Hipotesis yang digunakan untuk uji parsial adalah:

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0, \text{ dengan } j = 1, 2, \dots, p. \end{aligned} \tag{11}$$

Statistik uji untuk menguji hipotesis dari Persamaan (11) adalah uji *Wald* yang dinyatakan sebagai berikut (Pawitan, 2001):

$$W = \frac{\hat{\beta}_j}{\sqrt{\hat{V}\hat{a}r(\hat{\beta}_j)}} \tag{12}$$

dimana $\hat{V}\hat{a}r(\hat{\beta}_j)$ diperoleh dari elemen diagonal ke $(p + 1)$ dari matriks varians kovarians $Cov(\hat{\beta})$, dimana $Cov(\hat{\beta}) = [I(\hat{\beta})]^{-1} = -[H(\hat{\beta})]^{-1} \cdot I(\hat{\beta})^{-1}$ dan $H(\hat{\beta})^{-1}$ adalah berturut-turut menyatakan matriks Informasi Fisher dan matriks Hessian. Daerah kritis pengujian hipotesis pada Persamaan (12) untuk tingkat signifikansi α adalah tolak H_0 jika nilai $|W| > Z_{\alpha/2}$ atau tolak H_0 bila *p-value* kurang dari α . Nilai $Z_{\alpha/2}$ dapat diperoleh dari tabel distribusi normal standar.

Klasifikasi yang berasal dari variabel respon biner dilakukan dengan cara menentukan nilai titik potong. Titik potong yang dapat digunakan yaitu sebesar 0.5. Ketentuan klasifikasi yang digunakan adalah sebagai berikut:

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

Apabila probabilitas yang dihasilkan dari model bernilai lebih kecil dari 0,5 maka hasil prediksi adalah kategori 0, sedangkan peluang yang dihasilkan dari model bernilai lebih besar atau sama dengan 0,5 maka hasil prediksi adalah kategori 1 (Bishop, 2006).

2.2 Uji Independensi

Salah satu asumsi yang harus dipenuhi dalam analisis regresi dengan beberapa variabel prediktor adalah tidak terdapat hubungan antara satu variabel prediktor dengan variabel prediktor yang lain. Adanya hubungan antara dua variabel prediktor dalam model regresi menyebabkan taksiran parameter model regresi yang dihasilkan akan memiliki *error* yang sangat besar. Misalkan A_1 dan A_2 adalah dua variabel prediktor yang akan diuji. Hipotesis ujinya adalah:

$$\begin{aligned} H_0 &: \text{ Variabel } A_1 \text{ dan } A_2 \text{ saling independen} \\ H_1 &: \text{ Variabel } A_1 \text{ dan } A_2 \text{ saling dependen} \end{aligned}$$

Statistik uji yang digunakan dalam pengujian ini menggunakan statistik *Pearson chi-square* yang ditunjukkan pada Persamaan (13) yang diformulasikan sebagai berikut :

$$\chi^2 = \sum_{b=1}^B \sum_{k=1}^K \frac{(O_{bk} - E_{bk})^2}{E_{bk}} \tag{13}$$

dimana :

O_{bk} : Banyaknya observasi yang termasuk pada variabel A_1 kategori ke- b dan variabel A_2 kategori ke- k .
 E_{bk} : Nilai harapan untuk observasi yang termasuk pada variabel A_1 kategori ke- b dan variabel A_2 kategori ke- k . Nilai E_{bk} dihitung menggunakan rumus

$$E_{bk} = \frac{n_b \cdot n_k}{n}, \quad \begin{aligned} b &= 1, 2, \dots, B \\ k &= 1, 2, \dots, K \end{aligned}$$

B, K : Banyaknya kategori untuk masing-masing variabel A_1 dan A_2 .

Statistik *Pearson chi-square* pada Persamaan (13) berdistribusi *chi-square* dengan derajat bebas, $db = (B - 1)(K - 1)$ (Bilder and Laughin, 2015). Oleh karena itu, daerah penolakan H_0 dari pengujian hipotesis pada persamaan (2.13) untuk tingkat signifikansi α adalah tolak H_0 jika nilai $\chi^2 > \chi^2_{(\alpha, db)}$ atau tolak H_0 bila *p-value* kurang dari α . Nilai $\chi^2_{(\alpha, db)}$ didapat dari tabel distribusi *chi-square*.

2.3 Neural Networks

Menurut Jaya, dkk (2018), pembagian arsitektur *neural networks* dapat dilihat dari kerangka kerja dan skema interkoneksi. Kerangka kerja *neural networks* dapat dilihat dari jumlah lapisan (*layer*) dan jumlah node pada setiap lapisan. Lapisan-lapisan penyusun *neural networks* dapat dibagi menjadi tiga, yaitu:

1. Lapisan *input*
2. Lapisan tersembunyi (*hidden layer*)
3. Lapisan *output*

Backpropagation

Backpropagation atau propagasi balik merupakan suatu metode pembelajaran atau pelatihan *supervised learning* yang banyak digunakan. Pada jaringan *backpropagation*, setiap unit yang berada di lapisan *input*

terhubung dengan setiap unit yang ada di *hidden layer*. Setiap unit yang ada di *hidden layer* terhubung dengan setiap unit yang ada di lapisan *output*. Jaringan ini terdiri dari banyak lapisan (*multilayer net*) (Puspitaningrum, 2006).

Algoritma Backpropagation

Menurut Pramana dkk (2018), asumsi yang digunakan algoritma *backpropagation* yaitu fungsi aktivasi sigmoid biner. Langkah-langkah algoritma *backpropagation* sebagai berikut:

1. Menentukan *learning rate*, jumlah *hidden layer*, target *error* dan batas iterasi.
2. Inisialisasi nilai bobot w dan v .

Tahap Feed-forward

3. Masing-masing unit *input* ($x_k, k = 1, 2, \dots, p$) menerima sinyal *input* x_k dan menjalankan sinyal tersebut ke setiap unit pada *hidden layer*.
4. Untuk setiap unit *hidden layer* ($z_m, m = 1, 2, \dots, r$) dijumlahkan bobotnya dengan sinyal *input* masing-masing:

$$z_{in_m} = v_{0m} + \sum_{k=1}^p x_k v_{km} \tag{14}$$

Menerapkan fungsi aktivasi untuk menghitung nilai sinyal *hidden layer* :

$$z_m = f(z_{in_m}) \tag{15}$$

Kemudian sinyal ini dikirimkan ke semua unit lapisan atasnya.

5. Untuk setiap unit *output* ($y_i, i = 1, 2, \dots, n$) dijumlahkan bobotnya dengan sinyal *input* masing-masing:

$$y_{in_i} = w_{0i} + \sum_{m=1}^r z_m w_{mi} \tag{16}$$

Menerapkan fungsi aktivasi untuk menghitung nilai sinyal *output* :

$$y_i = f(y_{in_i}) \tag{17}$$

Tahap Backpropagation of Error

6. Untuk setiap unit *output* ($y_i, i = 1, 2, \dots, n$) menerima sebuah pola target yang berhubungan dengan pola *input* pelatihan, kemudian menghitung nilai *error*:

$$\delta_i = (T_i - y_i)y_i(1 - y_i) \tag{18}$$

Menghitung perubahan bobotnya

$$\Delta w_{mi} = \alpha \delta_i z_m \tag{19}$$

Menghitung perubahan bias

$$\Delta w_{0i} = \alpha \delta_i \tag{20}$$

7. Menghitung *error* sinyal *output* ($z_m, m = 1, 2, \dots, r$)

$$\delta_{net_m} = \delta_i w_{mi} \tag{21}$$

Menghitung nilai *error*

$$\delta_m = \delta_{net_m} z_m (1 - z_m) \tag{22}$$

Menghitung perubahan bobotnya

$$\Delta v_{km} = \alpha \delta_m x_k \tag{23}$$

Kemudian menghitung perubahan biasnya

$$\Delta v_{0m} = \alpha \delta_m \tag{24}$$

Tahap Update Bobot dan Bias

8. Untuk setiap unit *output* dilakukan *update* bobot dan biasnya

$$w_{mi(\text{baru})} = w_{mi} + \Delta w_{mi} \tag{25}$$

Untuk setiap unit *hidden layer* dilakukan *update* bobot dan biasnya

$$v_{km(\text{baru})} = v_{km} + \Delta v_{km} \tag{26}$$

9. Uji/periksa kondisi berhenti (akhir iterasi). Kondisi berhenti ini terpenuhi apabila nilai *error* yang dihasilkan lebih kecil dari target *error* yang ditetapkan. Selama kondisi berhenti masih tidak terpenuhi, lakukan langkah 3 sampai dengan 9.
10. Kemudian dilakukan klasifikasi menggunakan tahap *feed-forward*. Berdasarkan aturan metode klasifikasi *neural networks*, jika $0 \leq |y_i - \hat{y}_i| < 0,5$ maka dikatakan tepat klasifikasi.

2.4 Confusion matrix

Confusion matrix merupakan sebuah tabel yang membantu untuk menilai seberapa baik sebuah *classifier*. Tabel *confusion matrix* dihasilkan dari aplikasi model pada *test set*. Tabel ini merupakan tabel *binary*, dimana kelas dibagi menjadi 2 label kelas. Format dari *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Actual Class	Predicted Class	
	Kelas = 0	Kelas = 1
Kelas = 0	TP	FN
Kelas = 1	FP	TN

Keterangan dari Tabel 1, sebagai berikut:

Actual Class = Kelas yang sebenarnya pada *test set*

Predicted Class = Kelas hasil prediksi dari model yang dihasilkan oleh *classifier*

True Positive (TP) = Jumlah baris berlabel kelas 0 pada *test set* yang benar diklasifikasikan sebagai kelas 0 oleh *classifier*

False Negative (FN) = Jumlah baris berlabel kelas 0 pada *test set*, namun diklasifikasikan sebagai kelas 1 oleh *classifier*

False Positive (FP) = Jumlah baris berlabel kelas 1 pada *test set*, namun diklasifikasikan sebagai kelas 0 oleh *classifier*

True Negative (TN) = Jumlah baris berlabel kelas 1 pada *test set* yang benar diklasifikasikan sebagai kelas 1 oleh *classifier* [14].

Berdasarkan Tabel 1, perhitungan akurasi dapat dinyatakan pada Persamaan (27). Akurasi tersebut merupakan persentase baris *test set* yang diklasifikasikan dengan benar (Pramana dkk, 2018).

$$Akurasi = \frac{(TP + TN)}{TP + FN + FP + TN} \tag{27}$$

3. Metodologi Penelitian

3.1. Sumber Data dan Variabel Penelitian

Penelitian ini dilakukan di Laboratorium Statistika Terapan dan Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman dari Desember 2021 sampai dengan Maret 2024. Data yang digunakan dalam penelitian ini data lulusan program sarjana FMIPA Universitas Mulawarman, dengan periode wisuda tahun 2021 sebanyak 188 lulusan. Variabel yang digunakan dalam penelitian ini:

- Y : Status Predikat Lulusan
- X₁ : Jenis Kelamin
- X₂ : Asal Daerah
- X₃ : Jalur Penerimaan
- X₄ : Status SMA/SMK/MA
- X₅ : Program Studi.

3.2. Metode Analisis

Adapun langkah-langkah untuk mencapai tujuan pada penelitian ini adalah sebagai berikut:

- 1) Pengacakan dan pembagian proporsi data
- 2) Klasifikasi menggunakan Regresi Logistik Biner, dengan tahapan sebagai berikut:
 - a. Melakukan uji independensi *Pearson chi-square* menggunakan Persamaan (13).
 - b. Membentuk model awal regresi logistik biner dengan menggunakan data *training*.
 - c. Menguji signifikansi parameter secara serentak menggunakan Persamaan (10).
 - d. Menguji parameter secara parsial, untuk mengetahui variabel-variabel prediktor yang mempengaruhi variabel respon secara parsial menggunakan Persamaan (12).
 - e. Membentuk model akhir regresi logistik biner.
 - f. Menentukan klasifikasi data *testing* menggunakan model akhir.
 - g. Menghitung nilai akurasi pada hasil klasifikasi dengan metode regresi logistik biner menggunakan Persamaan (27).
- 3) Klasifikasi dengan metode *neural networks*, dengan tahapan sebagai berikut:
 - a. Menentukan *learning rate*, jumlah *hidden layer* target *error* dan batas iterasi. Menggunakan nilai *learning rate* yang sebesar 0,01; jumlah *hidden layer* sebanyak 5; target *error* sebesar 0,001; dan batas iterasi sebanyak 500. Fungsi aktivasi yang digunakan adalah sigmoid biner.
 - b. Menentukan bobot dan bias.
 - c. Melakukan pelatihan *backpropagation* dengan data *training*
Proses *feed-forward*
 1. Menjumlahkan bobot dan bias sinyal *input* pada setiap unit *hidden layer* menggunakan Persamaan (14).
 2. Menggunakan fungsi aktivasi sigmoid biner untuk menghitung sinyal *hidden layer* berdasarkan Persamaan (15).

3. Menjumlahkan bobot dan bias sinyal *input* pada setiap unit *output* menggunakan Persamaan (16).
4. Menggunakan fungsi aktivasi sigmoid biner untuk menghitung sinyal *output* berdasarkan Persamaan (17).
Proses *backpropagation*
5. Menghitung *error* dari *output* menggunakan Persamaan (18).
6. Menghitung suku koreksi bobot dan bias untuk menentukan w_{mi} dan w_{oi} menggunakan Persamaan (19) dan (20).
7. Menghitung *error* sinyal *output* pada *hidden layer* menggunakan Persamaan (21) dan (22).
8. Menghitung suku koreksi bobot dan bias untuk menentukan v_{km} dan v_{om} menggunakan Persamaan (23) dan (24).
Proses *update* bobot dan bias
9. Menentukan bobot dan bias baru pada setiap *output* menggunakan Persamaan (25).
10. Menentukan bobot dan bias baru pada setiap *hidden layer* menggunakan Persamaan (26).
- d. Mengulangi langkah-langkah pada poin c untuk data *training* kedua sampai terakhir dengan bobot dan bias baru. Kondisi berhenti apabila *error* telah terpenuhi.
- e. Setelah mendapatkan bobot dan bias terbaik, kemudian dilakukan pengujian pada data *testing* dengan proses *feed-forward* untuk mendapatkan hasil klasifikasi.
- f. Menghitung nilai akurasi pada hasil klasifikasi dengan metode *neural networks* menggunakan Persamaan (27).
- 4) Membandingkan nilai akurasi dari hasil klasifikasi antara metode regresi logistik biner dan *neural networks*, untuk mengetahui metode yang lebih baik.
- 5) Mengambil kesimpulan dari perbandingan nilai akurasi antara metode regresi logistik biner dan *neural networks*.

4. Hasil dan Pembahasan

1) Klasifikasi Status Predikat Lulusan Sarjana Menggunakan Regresi Logistik Biner

Data yang digunakan merupakan data lulusan program sarjana FMIPA Universitas Mulawarman Tahun 2021. Dalam penelitian ini digunakan proporsi data *training* adalah 70% sehingga data urutan pertama sampai ke 132 hasil pengacakan akan dijadikan sebagai data *training*. Kemudian 30% data sisanya yakni urutan 133 sampai urutan 188 akan menjadi data *testing*.

Uji Independensi Antar Variabel Prediktor

Salah satu asumsi yang harus dipenuhi dalam analisis regresi dengan beberapa variabel prediktor adalah tidak terjadinya korelasi antara satu variabel prediktor dengan variabel prediktor yang lain (tidak terjadi multikolinieritas). Perhitungan uji independensi antar variabel prediktor dapat dilihat pada Tabel 2.

Tabel 2. Hasil Uji Multikolinieritas Variabel Prediktor

Variabel Prediktor	Statistik Pearson Chi-square	p-value	Keputusan	Kesimpulan
X_1 dan X_2	2,44	0,11*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_1 dan X_3	3,59	0,17*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_1 dan X_4	0,84	0,36*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_1 dan X_5	0,87	0,83*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_2 dan X_3	3,16	0,21*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_2 dan X_4	2,57	0,11*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_2 dan X_5	3,03	0,39*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_3 dan X_4	5,09	0,08*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_3 dan X_5	8,84	0,18*	H_0 gagal ditolak	Tidak terjadi multikolinieritas
X_4 dan X_5	4,48	0,21*	H_0 gagal ditolak	Tidak terjadi multikolinieritas

Berdasarkan Tabel 2 dapat diketahui bahwa tidak terjadi multikolinieritas pada seluruh hasil pengujian antar variabel prediktor. Sehingga seluruh variabel prediktor dapat dilanjutkan dengan metode regresi logistik biner.

Pemodelan Status Predikat Lulusan Program Sarjana dengan Regresi Logistik Biner

Pemodelan status predikat lulusan program sarjana FMIPA Universitas Mulawarman dengan regresi logistik biner diawali dengan penaksiran parameter model regresi logistik biner. Penaksiran parameter model regresi logistik biner dengan metode MLE dan iterasi *Fisher Scoring* didapatkan hasil seperti pada Tabel 3.

Tabel 3. Taksiran Parameter Regresi Logistik Biner

Parameter	Taksiran Parameter
$\hat{\beta}_0$	4,1484
$\hat{\beta}_{1(1)}$	0,0013
$\hat{\beta}_{2(1)}$	0,0248
$\hat{\beta}_{3(1)}$	-0,8775
$\hat{\beta}_{3(2)}$	-2,5963
$\hat{\beta}_{4(1)}$	16,1918
$\hat{\beta}_{5(1)}$	-0,0235
$\hat{\beta}_{5(2)}$	-1,4251
$\hat{\beta}_{5(3)}$	-0,3056

Berdasarkan nilai taksiran parameter model regresi logistik biner pada Tabel 3, maka diperoleh model regresi sebagai berikut:

$$g(x_i) = 4,1484 + 0,0013x_{1(1)} + 0,0248x_{2(1)} - 0,8775x_{3(1)} - 2,5963x_{3(2)} + 16,1918x_{4(1)} - 0,0235x_{5(1)} - 1,4251x_{5(2)} - 0,3056x_{5(3)} \tag{28}$$

dengan
$$\pi(x_i) = \frac{\exp g(x_i)}{1 + \exp g(x_i)}$$

dengan :

- $x_{1(1)}$: 1, untuk jenis kelamin lulusan program sarjana adalah perempuan dan 0, untuk lainnya.
- $x_{2(1)}$: 1, untuk lulusan program sarjana berasal dari luar Samarinda dan 0, untuk lainnya
- $x_{3(1)}$: 1, untuk jalur penerimaan lulusan program sarjana adalah SBMPTN dan 0, untuk lainnya
- $x_{3(2)}$: 1, untuk jalur penerimaan lulusan program sarjana adalah SMMPTN dan 0, untuk lainnya
- $x_{4(1)}$: 1, untuk status SMA/SMK/MA lulusan program sarjana adalah swasta dan 0, untuk lainnya
- $x_{5(1)}$: 1, untuk program studi lulusan program sarjana adalah program studi biologi dan 0, untuk lainnya
- $x_{5(2)}$: 1, untuk program studi lulusan program sarjana adalah program studi kimia dan 0, untuk lainnya
- $x_{5(3)}$: 1, untuk program studi lulusan program sarjana adalah program studi fisika dan 0, untuk lainnya.

Hasil Klasifikasi dengan Metode Regresi Logistik Biner

Selanjutnya, didapatkan kesimpulan hasil klasifikasi metode regresi logistik biner pada data *testing* yang disajikan pada Tabel 4.

Tabel 4. *Confusion Matrix* Metode Regresi Logistik Biner pada Data *Testing*

Klasifikasi Status Predikat Lulusan Program Sarjana FMIPA	Prediksi Klasifikasi Metode Regresi Logistik Biner		Total
	Cukup dan Memuaskan	Sangat Memuaskan dan Dengan Pujian	
Cukup dan Memuaskan	0	7	7
Sangat Memuaskan dan Dengan Pujian	0	49	49
Total	0	56	56

Berdasarkan Tabel 4, dapat diketahui bahwa status predikat lulusan program sarjana FMIPA Universitas Mulawarman Tahun 2021 dengan menggunakan metode regresi logistik biner diperoleh hasil yaitu 7 yang tidak tepat diklasifikasikan pada status predikat cukup dan memuaskan. Terdapat 49 yang tepat diklasifikasikan pada status predikat sangat memuaskan dan dengan pujian. Sehingga berdasarkan Persamaan (27) maka diperoleh nilai akurasi sebagai berikut :

$$\begin{aligned} Akurasi &= \frac{(0 + 49)}{(0 + 7 + 0 + 49)} \times 100\% \\ &= \frac{49}{56} \times 100\% \\ &= 87,5\%. \end{aligned}$$

2) Klasifikasi Status Predikat Lulusan Sarjana Menggunakan Neural Networks
Pemilihan Model *Neural Networks* Terbaik

Tabel 5 Nilai *Error* untuk Setiap *Hidden Neuron*

<i>Hidden Neuron</i>	<i>Error</i>
<i>Neural Networks-1</i>	4,28108
<i>Neural Networks-2</i>	4,277011
<i>Neural Networks-3</i>	2,400374
<i>Neural Networks-4</i>	2,411605
<i>Neural Networks-5</i>	2,395247

Hasil pemodelan pada Tabel 5, menunjukkan arsitektur dengan *hidden neuron* 5 adalah arsitektur yang mempunyai kinerja lebih baik karena menghasilkan *error* yang terkecil yaitu sebesar 2,395247.

Perhitungan *Neural Networks* pada kelompok arsitektur 5 *hidden layer*

- a. Pada penelitian ini menggunakan nilai *learning rate* sebesar 0,01; jumlah *hidden layer* sebanyak 5; target *error* sebesar 0,001; dan fungsi aktivasi yang digunakan adalah sigmoid biner. Selanjutnya dilakukan tahapan perhitungan dan didapatkan bobot terbaik untuk 5 *hidden layer* yang digunakan dalam mengklasifikasi data *testing*. Bobot dan bias tersebut adalah sebagai berikut:

$$v' = \begin{bmatrix} -2,8295 & -3,6047 & 2,0539 & -1,0154 & 0,9939 \\ -2,3910 & 1,1443 & -3,6912 & -1,9252 & 1,0896 \\ 9,4174 & 0,7814 & 3,6622 & 0,1907 & -1,1314 \\ -1,6282 & 1,8886 & 0,1444 & 0,2645 & -1,4190 \\ -4,9439 & -3,4705 & 1,9376 & -0,6190 & 1,6709 \\ -2,6648 & 5,9282 & -1,6524 & -1,0357 & -1,9620 \end{bmatrix}$$

$$w' = \begin{bmatrix} -11,2796 \\ 8,4105 \\ 6,0144 \\ 0,3086 \\ -2,2246 \\ 0,5336 \end{bmatrix}$$

- b. Setelah mendapatkan bobot dan bias terbaik, kemudian dilakukan pengujian pada data *testing* dengan proses *feed-forward*. Pada data *testing* pertama diketahui nilai $X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1$, dan $Y = 1$. Tahapan terakhir yaitu perhitungan dengan menggunakan fungsi aktivasi sigmoid biner untuk menghitung sinyal *output*, dan diperoleh

$$y_1 = \frac{1}{1 + e^{-y_{in_1}}} = 0,9831.$$

Sehingga dapat diketahui nilai *output* pada data *testing* pertama sebesar 0,9831. Berdasarkan aturan metode klasifikasi *neural networks*, jika $0 \leq |y_i - \hat{y}_i| < 0,5$ maka dikatakan tepat klasifikasi. Pada data *testing* pertama diketahui nilai $y_i = 1$ dan $\hat{y}_i = 0,9831$ maka obyek data pertama $0 \leq |0,0169| < 0,5$ tepat klasifikasi. Berdasarkan hasil klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menggunakan *neural networks* dengan 56 data *testing*.

Hasil Klasifikasi dengan Metode *Neural Networks*

Selanjutnya, didapatkan kesimpulan hasil klasifikasi metode *neural networks* pada data *testing* yang disajikan pada Tabel 6.

Tabel 6 *Confusion Matrix* Metode *Neural Networks* pada Data *Testing*

Klasifikasi Status Predikat Lulusan Program Sarjana FMIPA	Prediksi Klasifikasi Metode <i>Neural Networks</i>		Total
	Cukup dan Memuaskan	Sangat Memuaskan dan Dengan Pujian	
Cukup dan Memuaskan	0	7	7
Sangat Memuaskan dan Dengan Pujian	0	49	49
Total	0	56	56

Berdasarkan Tabel 6 dapat diketahui bahwa status predikat lulusan program sarjana FMIPA Universitas Mulawarman Tahun 2021 dengan menggunakan metode regresi logistik biner diperoleh hasil yaitu 7 yang tidak tepat diklasifikasikan pada status predikat cukup dan memuaskan. Terdapat 49 yang tepat diklasifikasikan pada status predikat sangat memuaskan dan dengan pujian. Sehingga berdasarkan Persamaan (27) maka diperoleh nilai akurasi sebagai berikut:

$$Akurasi = \frac{(0 + 49)}{(0 + 7 + 0 + 49)} \times 100\%$$

$$\begin{aligned}
 &= \frac{49}{56} \times 100\% \\
 &= 87,5\%.
 \end{aligned}$$

3) Perbandingan Nilai Akurasi dengan Metode Regresi Logistik Biner dan *Neural Networks*

Setelah proses klasifikasi menggunakan metode regresi logistik biner dan *neural networks* dilakukan pada data *testing*, maka diperoleh hasil persentase akurasi pada kedua metode tersebut yang dapat ditampilkan pada Tabel 7.

Tabel 7. Nilai Akurasi Metode Regresi Logistik Biner dan *Neural Networks*

Metode	Akurasi
Regresi Logistik Biner	87,5%
<i>Neural Networks</i>	87,5%

Berdasarkan Tabel 7, dapat diambil kesimpulan bahwa metode regresi logistik biner dan *neural networks* memiliki nilai akurasi yang sama dalam memprediksi klasifikasi yaitu sebesar 87,5%. Hal ini menunjukkan bahwa pada kedua metode tersebut pada kasus klasifikasi status predikat lulusan FMIPA Tahun 2021 memiliki tingkat yang sama baiknya.

5. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

- 1) Hasil klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menggunakan regresi logistik biner adalah terdapat 7 lulusan program sarjana FMIPA Universitas Mulawarman dengan status predikat cukup dan memuaskan yang tidak tepat diklasifikasikan. Terdapat 49 lulusan program sarjana FMIPA Universitas Mulawarman dengan status predikat memuaskan dan dengan pujian yang tepat diklasifikasikan.
- 2) Hasil klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menggunakan *neural networks* adalah terdapat 7 lulusan program sarjana FMIPA Universitas Mulawarman dengan status predikat cukup dan memuaskan yang tidak tepat diklasifikasikan. Terdapat 49 lulusan program sarjana FMIPA Universitas Mulawarman dengan status predikat memuaskan dan dengan pujian yang tepat diklasifikasikan.
- 3) Hasil perbandingan nilai akurasi antara regresi logistik biner dan *neural networks* pada klasifikasi status predikat lulusan program sarjana FMIPA Universitas Mulawarman menunjukkan nilai akurasi dalam memprediksi klasifikasi dengan persentase yang sama yaitu sebesar 87,5%. Hal ini menunjukkan bahwa tingkat akurasi klasifikasi pengklasifikasian metode regresi logistik ordinal sama dengan metode *neural networks*.

6. Daftar Pustaka

Agresti, A. (2019). *An Intoruction To Categorical Data Analysis Third Edition*. Florida: John Wiley & Sons, Inc.

Aji, C. M. (2014). Analisis Faktor-faktor yang Memengaruhi Laju Pertumbuhan Penduduk Kota Semarang Tahun 2011 Menggunakan Geographically Weighted Logistic Regression. *Jurnal Gaussian*, 3(2), 161-171.

Bilder, C.R., and Loughin, T.M. (2015). *Analysis of Categorical Data with R*. Boca Raton: CRC Press.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media.

Fatonah, L., Sanapiah, S., & Febrilia, B. R. (2019). Regresi Logistik Ordinal (Studi Kasus Faktor yang Mempengaruhi Tingkat Stres Mahasiswa dalam Menyelesaikan Skripsi). *Media Pendidikan Matematika*, 5(2), 146.

Fathurahman, M. (2017). Regresi Logistik untuk Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Pulau Kalimantan. *Seminar Nasional Matematika dan Aplikasinya*, ISBN 978-602-14413-1-2, 285.

Hasanah, S. H. (2019). Perbandingan Metode Klasifikasi Artificial Neural Network Backpropagation dan Regresi Logistik (Studi Kasus : Bank Internasional Indonesia). *Jurnal Statistika dan Matematika*, 1(1), 44-62.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression Third Edition*. New York: John Wiley & Sons, Inc.

Imaslihkah, S., Ratna, M., & Ratnasari, V. (2013). Analisis Regresi Logistik Ordinal Terhadap Faktor-Faktor

- yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya. *Jurnal Sains dan Seni Pomits*, 2(2), 177-182.
- Jaya, H., Sabran, Indris, M. M., Djawad, Y. A., Ilham, A., & Ahmar, A. S. (2018). *Kecerdasan Buatan*. Makassar: Universitas Negeri Makassar.
- Menezes, F. S., Liska, G. R., Cirillo, M. A., & Vivanco, M. J. (2017). Data Classification with Binary Response Through The Boosting Algorithm and Logistic Regression. *Expert Systems With Application*, 69(1), 63-65.
- Nurrohmat, M. A., & Nugroho, Y. S. (2015). Aplikasi Pemrediksi Masa Studi dan Predikat. *Khazanah Informatika*, 1(1), 29-34.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Oxford University Press.
- Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R. (2018). *Data Mining dengan R Konsep Serta Implementasi*. Bogor: In Media.
- Puspitaningrum, D. (2006). *Pengantar Jaringan Saraf Tiruan*. Yogyakarta: Penerbit Andi.
- Ramli, Yuniarti, D., & Goejantoro, R. (2013). Perbandingan Metode Klasifikasi Regresi Logistik dengan Jaringan Saraf Tiruan (Studi Kasus: Pemilihan Jurusan Bahasa dan IPS Pada SMAN 2 Samarinda Tahun Ajaran 2011/2012). *Jurnal Ekspansional*, 4(1), 17-24.
- Sutojo, T., Mulyanto, E., & Suharto, V. (2011). *Kecerdasan Buatan*. Yogyakarta: Penerbit Andi.
- Varamita, A. (2017). *Analisis Regresi Logistik dan Aplikasinya pada Penyakit Anemia untuk Ibu Hamil di RSKD Ibu dan Anak Siti Fatimah makassar*. Makassar: Universitas Negeri Makassar.