

Komparasi Perceptron dan Regresi Logistik pada Klasifikasi Data Breast Cancer Wisconsin

^{1,*}Sandi Badiwibowo Atim, ²Muhammad Afdhaluddin

^{1,2} Jurusan Ilmu Komputer, FMIPA, Universitas Lampung, Bandar Lampung, Indonesia

Abstrak — Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap dua model klasifikasi dasar dalam *machine learning*, yaitu Perceptron klasik dan Regresi Logistik, pada tugas klasifikasi data medis menggunakan dataset *Breast Cancer Wisconsin (Diagnostic)*. Dataset yang digunakan terdiri dari 569 sampel dengan 30 fitur numerik yang merepresentasikan karakteristik morfologi sel kanker hasil biopsi, seperti radius, tekstur, perimeter, area, dan *smoothness*. Penelitian ini diawali dengan tahap pra-pemrosesan data berupa normalisasi menggunakan metode *Min-Max Scaling* untuk memastikan setiap fitur berada pada rentang skala yang sama sehingga tidak mendominasi proses pembelajaran model. Selanjutnya, data dibagi menggunakan teknik *Hold-Out* dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Proses pelatihan dilakukan pada kedua model dengan parameter standar tanpa optimasi lanjutan untuk menjaga objektivitas perbandingan. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* guna memberikan gambaran menyeluruh terhadap performa klasifikasi, khususnya dalam konteks data medis yang sensitif terhadap kesalahan prediksi. Hasil penelitian menunjukkan bahwa Regresi Logistik memperoleh performa lebih tinggi dengan akurasi sebesar 96,49%, dibandingkan Perceptron klasik yang mencapai 93%. Temuan ini mengindikasikan bahwa pendekatan probabilistik pada Regresi Logistik lebih efektif dalam memodelkan hubungan antar fitur numerik dan menghasilkan batas keputusan yang lebih stabil pada klasifikasi kanker payudara. Sekaligus menegaskan kontribusi utama penelitian ini dalam menyajikan analisis komparatif terhadap kinerja model Perceptron dan Regresi Logistik dalam menangani data yang tidak sepenuhnya terpisah secara linier, serta memberikan gambaran praktis mengenai efektivitas penggunaan model linier sederhana berdasarkan hasil evaluasi kinerja klasifikasi pada dataset medis.

Kata Kunci: *breast cancer; klasifikasi; logistic regression; machine learning; perceptron;*

Abstract — This study aims to conduct a comparative analysis of two fundamental classification models in machine learning, namely the classical Perceptron and Logistic Regression, for medical data classification using the Breast Cancer Wisconsin (Diagnostic) dataset. The dataset consists of 569 samples with 30 numerical features representing morphological characteristics of cancer cells obtained from biopsy results, including radius, texture, perimeter, area, and smoothness. The research begins with a data preprocessing stage involving normalization using the Min-Max Scaling method to ensure that all features are within the same range and do not disproportionately influence the learning process. The dataset is then split using the Hold-Out technique with an 80:20 ratio for training and testing sets, respectively. Both models are trained using standard parameter settings without additional hyperparameter optimization to maintain fairness in comparison. Model performance is evaluated using accuracy, precision, recall, and F1-score to provide a comprehensive assessment, particularly in the context of medical data where misclassification can have critical consequences. The results indicate that Logistic Regression achieves higher performance with an accuracy of 96.49%, compared to 93% for the classical Perceptron. These findings suggest that the probabilistic approach of Logistic Regression is more effective in modeling relationships among numerical features and producing a more stable decision boundary for breast cancer classification tasks, while also highlighting the main contribution of this study in presenting a comparative analysis of the performance of Perceptron and Logistic Regression models in handling data that are not perfectly linearly separable, as well as providing practical insights into the effectiveness of simple linear models based on classification performance evaluation on medical datasets.

Keywords: *breast cancer; classification; logistic regression; machine learning; perceptron*

* Corresponding author :
Sandi Badiwibowo Atim
Program Studi Manajemen Informatika, Bandar Lampung, Indonesia
sandibadiwibowoatim@fmipa.unila.ac.id

1. PENDAHULUAN

Dalam beberapa dekade terakhir, pembelajaran mesin (*machine learning*) telah menjadi pendekatan utama dalam pengembangan sistem cerdas, terutama dalam bidang klasifikasi data medis. Penerapan *machine learning* dapat mengotomasi proses pembuatan model analitis dan menyelesaikan tugas-tugas yang diberikan[1]. Salah satu cabang paling mendasar adalah klasifikasi biner, yang bertujuan mengelompokkan data ke dalam dua kelas berdasarkan pola dari data historis. Dalam konteks diagnosis medis, klasifikasi ini biasanya digunakan untuk memetakan fitur data guna menjawab pertanyaan 'ya' atau 'tidak' (misalnya, apakah pasien mengidap kanker atau tidak[2]). Berbagai algoritma telah dikembangkan, mulai dari model sederhana seperti Perceptron dan Regresi Logistik hingga model yang lebih kompleks.

Perceptron, yang diperkenalkan oleh Frank Rosenblatt pada 1958, merupakan algoritma jaringan syaraf tiruan pertama yang menyesuaikan bobot berdasarkan kesalahan output, namun memiliki keterbatasan hanya pada kasus linier. Perceptron merupakan algoritma yang paling sederhana pada Jaringan Saraf Tiruan (*Artificial Neural Network*) karena hanya memiliki lapisan tunggal (*single layer*) yang terdiri dari bobot (*weights*) dan bias untuk menghasilkan keluaran[3]. Keunggulan dari model ini adalah strukturnya yang sederhana, proses pelatihan yang cepat, dan kemampuannya mengklasifikasikan data yang bersifat terpisah secara linier (*linearly separable*)[4]. Sementara itu, Regresi Logistik muncul sebagai pendekatan linier yang lebih stabil dengan output probabilistik melalui fungsi aktivasi sigmoid[5]. Secara teknis, model ini menggunakan pendekatan linier yang lebih stabil dengan menghasilkan output probabilistik melalui fungsi aktivasi sigmoid, yang memetakan angka apa pun ke dalam rentang nilai antara 0 dan 1[6]. Regresi logistik menawarkan keseimbangan antara kinerja prediktif dan interpretabilitas yang baik untuk aplikasi medis dibandingkan model yang terlalu kompleks[7]. Diagnosis kanker payudara saat ini masih sangat bergantung pada interpretasi ahli patologi terhadap hasil pemeriksaan sitologi atau histopatologi yang bersifat subjektif dan dapat menghasilkan variasi diagnosis antar pemeriksa[7]. Penelitian ini memfokuskan pada perbandingan kinerja kedua model tersebut menggunakan dataset *Breast Cancer Wisconsin* untuk mengidentifikasi model yang paling efektif sebagai pendekatan awal dalam klasifikasi medis.

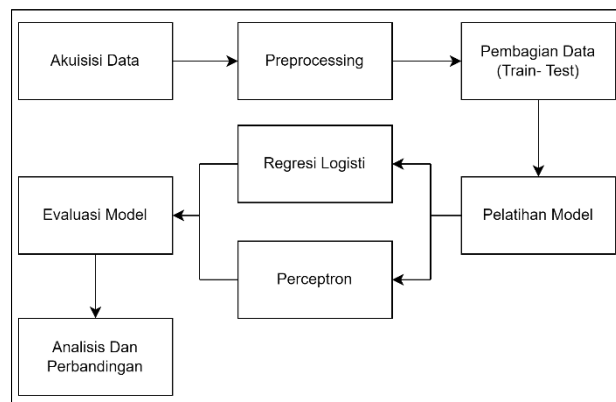
Beberapa penelitian sebelumnya telah membandingkan berbagai algoritma klasifikasi dalam domain medis. Penelitian oleh Sari et al. (2023) menunjukkan bahwa Multilayer Perceptron mampu mencapai akurasi tinggi dalam klasifikasi kanker payudara, namun dengan kompleksitas model yang lebih tinggi. Sementara itu, Khasanah et al. (2024) menunjukkan bahwa regresi logistik tetap menjadi metode yang stabil dan interpretable dalam prediksi data medis. Penelitian lain juga menunjukkan bahwa model neural network memiliki performa yang baik, namun membutuhkan proses pelatihan yang lebih kompleks serta tuning parameter yang tidak sederhana. Selain itu, penelitian oleh Wulandari dan Novita (2024) mengenai implementasi Feed Forward Neural Network (FFNN) dalam prediksi penyakit menunjukkan bahwa model berbasis neural network mampu memberikan performa yang tinggi, namun dengan kebutuhan komputasi dan kompleksitas yang lebih besar dibandingkan model linier. Hal ini mengindikasikan adanya trade-off antara akurasi dan kompleksitas model dalam penerapan machine learning pada data medis. Meskipun berbagai penelitian telah menunjukkan bahwa model berbasis neural network mampu mencapai performa klasifikasi yang tinggi, sebagian besar pendekatan tersebut melibatkan kompleksitas model yang tinggi serta kebutuhan komputasi yang besar. Di sisi lain, model linier seperti Regresi Logistik dikenal lebih stabil dan interpretable, namun masih jarang dibandingkan secara langsung dengan model sederhana seperti Perceptron klasik dalam kondisi tanpa optimasi parameter tambahan.

Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis komparatif antara Perceptron klasik dan Regresi Logistik pada dataset Breast Cancer Wisconsin guna mengevaluasi efektivitas kedua model dalam klasifikasi data medis. Adapun kontribusi utama dari penelitian ini adalah menyajikan perbandingan langsung antara dua model klasifikasi linier dasar, yaitu Perceptron dan Regresi Logistik, pada dataset yang sama tanpa penerapan optimasi parameter lanjutan; (2) mengevaluasi performa kedua model menggunakan metrik klasifikasi untuk memberikan gambaran objektif terhadap kemampuan

©2026 Ilmu Komputer Unila Publishing Network all rights reserved
 prediktif masing-masing metode; serta (3) memberikan analisis awal mengenai efektivitas penggunaan model sederhana dalam konteks klasifikasi medis sebagai alternatif pendekatan yang lebih kompleks. Hasil dari penelitian ini diharapkan dapat memberikan dasar pemahaman dalam pemilihan model klasifikasi yang efisien dan interpretable, khususnya pada aplikasi data medis dengan kompleksitas terbatas.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental. Tahapan metodologi mencakup akuisisi data, Preprocessing, Pembagian data (*Train-Test*), pelatihan model (Perceptron dan Regresi Linear), Evaluasi model lalu Analisis dan Perbandingan.



Gambar 1. Metode Penelitian

2.1. Akuisisi data

Data yang digunakan dalam penelitian ini adalah dataset *Breast Cancer Wisconsin (Diagnostic)* yang diperoleh dari sumber resmi UCI *Machine Learning Repository* melalui alamat repositori: <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data>. Dataset ini dipilih karena memiliki ukuran dan kompleksitas yang tepat untuk menguji performa model klasifikasi, namun tetap ringan untuk diolah, serta telah digunakan secara luas dalam berbagai literatur riset medis dan pembelajaran mesin. Dataset ini terdiri dari 569 sampel pasien yang diperoleh dari hasil biopsi jaringan payudara. Setiap sampel memiliki 30 fitur numerik yang menggambarkan karakteristik morfologi inti sel, seperti radius, tekstur, perimeter, area, dan kehalusan (*smoothness*). Target klasifikasi dalam data ini bersifat biner yang terbagi menjadi dua kelas, yaitu:

Tabel 1. Target Klasifikasi

Klasifikasi kanker	Jumlah Sampel	Persentase
<i>Malignant</i> (Ganas)	212	37,26%
<i>Benign</i> (Jinak)	357	62,74%
Total	569	100%

2.2. Preprocessing

Tahapan pra-pemrosesan data pada penelitian ini sangat krusial untuk memastikan stabilitas dari pelatihan model, terutama karena algoritma yang digunakan sangat sensitif terhadap skala input. *Preprocessing* dilakukan dengan tujuan utama untuk meningkatkan kualitas dan penyempurnaan data sebelum diproses oleh model, serta untuk meningkatkan akurasi perhitungan[8]. Langkah awal yang dilakukan adalah pelabelan data (*lable encoding*), di mana kolom diagnosis yang bersifat kategoris diubah menjadi numerik; kelas *Malignant* (M) diberi nilai 1 dan *Benign* (B) diberi nilai 0. Fitur-fitur input (X) mencakup 30 karakteristik morfologi sel, mulai dari *radius1* hingga *fractal_dimension3*. Selanjutnya dilakukan normalisasi fitur menggunakan metode *Min-Max Scaling* untuk menyamakan

bobot setiap fitur agar berada dalam rentang 0 hingga 1. Proses ini bertujuan untuk menghindari bias skala yang dapat memengaruhi perkalian bobot pada model Perceptron dan Regresi Logistik. Rumus normalisasi yang digunakan adalah :

$$x' = \frac{(x - \min(x_{\text{seluruh data}}))}{(\max(x_{\text{seluruh data}}) - \min(x_{\text{seluruh data}}))} \quad (1)$$

Tabel 2. Tabel Hasil Normalisasi

Radius1	Texture1	Parimeter1	Area1	Smoothness1	...	Fractal_dimension3
13.54	14.36	87.46	566.3	0.09779	...	0.07259
13.08	15.71	85.63	520	0.175	...	0.08183
...
20.6	29.33	140.1	1265	20.6		0.124

2.3. Pembagian dataset

Langkah terakhir dalam pra-pemrosesan adalah pembagian data (*data splitting*). Dataset dibagi menggunakan teknik *Hold-Out / Train-Test Split* dengan proporsi 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*). Data uji digunakan secara khusus untuk menilai kemampuan model dalam memprediksi data baru yang belum pernah dilihat sebelumnya[9]. Untuk menjaga agar distribusi kelas tetap seimbang antara data latih dan data uji, digunakan teknik *Stratified Sampling*. Pembagian ini dilakukan setelah data diacak (*shuffle*) guna menghindari *overfitting* serta memastikan generalisasi model dan evaluasi yang objektif[10].

2.4. Pelatihan model

Setelah tahapan pra-pemrosesan data selesai, langkah selanjutnya adalah melakukan proses pelatihan model untuk menemukan parameter optimal yang dapat membedakan antara *Malignan* dan *Benign*. Proses pelatihan ini melibatkan penyesuaian bobot (*weights*) dan bias secara iteratif berdasarkan data latih, di mana setiap model berusaha meminimalkan tingkat kesalahan prediksi melalui fungsi aktivasi masing-masing [11]. Penggunaan satu dataset yang sama bertujuan untuk membandingkan efektivitas model linier sederhana seperti Perceptron dengan pendekatan probabilistik pada Regresi Logistik dalam mengenali pola fitur medis yang kompleks[12].

2.4.1 Perceptron

Implementasi Perceptron klasik dimulai dengan memisahkan fitur (X) yang mencakup 30 kolom karakteristik sel dan label target (Y) yang telah diubah menjadi nilai biner. Inisialisasi parameter dilakukan dengan menetapkan nilai awal bobot dan bias sebesar 0, serta menggunakan *learning rate* (α) sebesar 0,1. Model ini menggunakan fungsi aktivasi step, di mana output akan bernilai 1 jika hasil kombinasi linier (z) lebih besar atau sama dengan 0, dan bernilai 0 jika sebaliknya. Selama proses pelatihan, dilakukan *forward pass* untuk menghitung prediksi dan dilanjutkan dengan pembaruan bobot menggunakan aturan delta jika terjadi kesalahan prediksi. Rumus pembaruan bobot yang digunakan adalah $w_i = w_i + \alpha * error * x_i$

epoch 1

Data 1 (output *malignant* atau 1)

Hitung z

$$z = w_i \cdot x_i + b$$

$$z = (0 \times 0.53192 + 0 \times 0.30402 + 0 \times 0.52871 + 0 \times 0.37731 + \dots + 0 \times 0.16286) + 0$$

$$z = 0$$

Prediksi

$y_pred = step(z) = z \geq 0$ maka return 1
 $y = 1$ = benar = tidak update bobot dan bias

Data 2 (output *benign* atau 0)

Hitung z

$$z = w_i \cdot x_i + b$$

$$z = (0 \times 0.19637 + 0 \times 0.4305 + 0 \times 0.19142 + 0 \times 0.09964 + \dots + 0 \times 0.16581) + 0$$

$$z = 0$$

Prediksi

$y_pred = step(z) = z \geq 0$ maka return 1

$y = 0$ -> salah -> update bobot dan *bias*!

$Error = y - y_pred = 0 - 1 = -1$

Update setiap bobot

$$w_i = w_i + \alpha * error * x_i$$

$$w_1 = 0 + 0.1 * -1 * 0.19637$$

$$w_1 = -0.019637$$

$$w_2 = -0.04305$$

....

$$bias = 0 + 0.1 * (-1) = -0.1$$

Semua langkah langkah diatas dilakukan terus menerus sampai semua data *training* selesai, dan ulangi kembali sampai $y_pred[i] = y$, dalam penelitian ini training mencapai 1807 *epoch* dan didapatkan nilai bobot *bias* dan *threshold* sebagai berikut

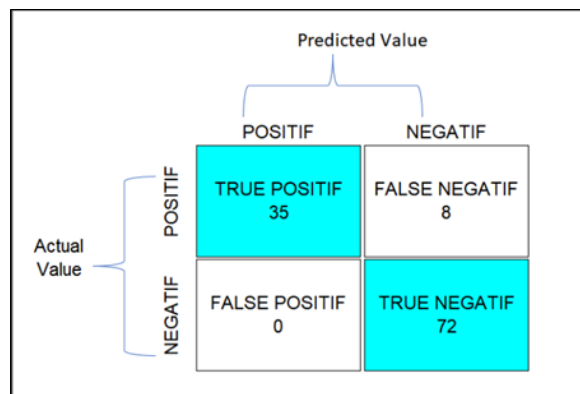
$W_1 = -1.88750$, $W_2 = 0.79976$, $W_3 = -1.25470$, $W_4 = -1.63443$, $W_5 = -1.69932$, $W_6 = -2.14355$, $W_7 = 0.94801$, $W_8 = 4.61151$, $W_9 = -0.74339$, $W_{10} = -0.16703$, $W_{11} = 4.67873$, $W_{12} = -0.16164$, $W_{13} = -1.23946$, $W_{14} = 4.69477$, $W_{15} = -2.57639$, $W_{16} = 0.70103$, $W_{17} = -2.01456$, $W_{18} = 4.26088$, $W_{19} = 0.15515$, $W_{20} = -5.64059$, $W_{21} = 2.37984$, $W_{22} = 1.28936$, $W_{23} = 2.61073$, $W_{24} = 3.97503$, $W_{25} = 2.28156$, $W_{26} = -1.09131$, $W_{27} = 0.79890$, $W_{28} = -2.07372$, $W_{29} = 0.98796$, $W_{30} = 0.98796$, $b = -2.60000$.

Setelah fase pelatihan mencapai 1807 epoch dan menghasilkan parameter bobot serta bias yang optimal, model Perceptron klasik diuji menggunakan data uji (*test set*) yang mencakup 20% dari total dataset. Proses pengujian ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Model melakukan prediksi dengan mengalikan fitur-fitur pada data uji dengan bobot akhir ($w_1 - w_{30}$) dan menjumlahkannya dengan nilai bias sebesar -2,6 . Hasil kombinasi linier tersebut kemudian diproses melalui fungsi aktivasi step, di mana nilai ≥ 0 diklasifikasikan sebagai 1 (*Malignant*) dan nilai < 0 sebagai 0 (*Benign*).

Tabel 3. Tabel Hasil Testing Perceptron

Data ke	Prediksi	Output yang diharapkan
1	1	1
2	0	0
3	0	0
...
21	0	1
22	1	1
26	0	1
39	1	1
41	0	1
51	0	0
67	0	1
98	0	1
101	0	1
108	0	1
115	0	1

Evaluasi Kinerja Perceptron Hasil pengujian pada data uji dievaluasi menggunakan *confusion matrix* untuk mendapatkan gambaran performa yang komprehensif.



Gambar 2. Confusion Matrix Perceptron

$$\begin{aligned}
 Accuracy &= (TP + TN) / (TP + TN + FP + FN) \\
 &= (35 + 72) / (35 + 72 + 0 + 8) = 0.93 \\
 Precision &= TP / (TP + FP) \\
 &= 35 / (35 + 0) = 1 \\
 Recall &= TP / (TP + FN) \\
 &= 35 / (35 + 8) = 0.81 \\
 F1\ Score &= 2 \times ((precision \times recall) / (precision + recall)) \\
 &= 2 \times ((1 \times 0.81) / (1 + 0.81)) = 0.894
 \end{aligned}$$

Berdasarkan perhitungan, model Perceptron klasik menghasilkan akurasi sebesar 93%, yang menunjukkan kemampuan klasifikasi yang cukup baik secara keseluruhan. Nilai *recall* (sensitivitas) model hanya mencapai 81%, yang mengindikasikan bahwa terdapat 8 kasus positif (*Malignant*) yang gagal teridentifikasi oleh model atau diklasifikasikan sebagai negatif palsu (*false negative*). Hal ini menyebabkan nilai *F1-score* berada pada angka 89,4%.

2.4.2 Regresi Logistik

Pada model Regresi Logistik, inialisasi bobot dilakukan secara acak dalam rentang kecil (-0,01 hingga 0,01) untuk menghindari masalah simetri agar model dapat mempelajari kompleksitas data secara efektif. Berbeda dengan Perceptron, model ini menggunakan fungsi aktivasi sigmoid. Fungsi aktivasi sigmoid adalah fungsi matematika dengan kurva berbentuk "S" yang mengubah output berupa probabilitas antara 0 dan 1[13]. Tahapan pelatihan meliputi *forward pass*, yaitu fase pertama dalam algoritma pelatihan untuk menghitung nilai kombinasi linier (z) dan probabilitas (y_{prob}) melalui rumus $y_{prob} = 1/(1+e^{-z})$ [14]. Evaluasi kesalahan selama pelatihan diukur menggunakan *Loss Function (Log Loss)* dimana *Loss Function* ini adalah komponen kunci dalam Regresi Logistik yang berfungsi untuk mengukur sejauh mana prediksi probabilitas model menyimpang dari label target sebenarnya selama fase pelatihan[14]. Kemudian diikuti dengan proses *Backpropagation* menggunakan *gradient descent*, metode ini berfungsi untuk mengubah bobot dalam arsitektur jaringan pelatihan yang digunakan untuk memperbaiki bobot[15]. Perbaikan bobot dilakukan dengan menghitung gradien terhadap setiap fitur guna meminimalkan total *mean loss*.

Data 1

$$\begin{aligned} \text{Hitung } z_1 \text{ (weighted sum)} &= z = w_1x_1 + w_2x_2 + \dots + w_{30}x_{30} + b \\ &= -0.0039 \times 0.27966 + -0.0072 \times 0.11532 + \dots + -0.0042 \times 0.09255 + 0 \\ &= -0.00375 \end{aligned}$$

σ = fungsi sigmoid

$$y_{prob} = \sigma(z) = 1 / (1 + e^{-z}) = 0.49829$$

$$loss = [y \log(\hat{y}) + (1 - y) * \log(1 - \hat{y})]$$

$$loss = -[0 \log(0.49829) + (1 - 0) * \log(1 - 0.49829)]$$

$$loss = 0.69165$$

$$\frac{\partial L}{\partial w_j} = (\hat{y} - y) * w_j$$

$$\frac{\partial L}{\partial w_j} = 1.78452$$

$$w_j = w_j - \eta \frac{\partial L}{\partial w_j}$$

$$w_j = -0.0090375$$

Data 2

$$\text{Hitung } z_2 \text{ (weighted sum)} = z_2 = \sum_{i=1}^n x_i + w_i + b$$

$$y_{prob} = \sigma(z) = 1 / (1 + e^{-0.000655}) = 0.49836$$

$$loss = -[0 \log(0.49836) + (1 - 0) \log(1 - 0.49836)]$$

$$loss = 0.69129$$

$$\frac{\partial L}{\partial w_j} = (\hat{y} - y) * w_j$$

$$\frac{\partial L}{\partial w_j} = 1.31970$$

$$w_j = w_j - \eta \frac{\partial L}{\partial w_j}$$

$$w_j = 0.054311$$

Pelatihan pada penelitian ini dihentikan pada *epoch* ke-100 karena model telah mencapai titik konvergen dengan nilai bias akhir sebesar -14,657.

Tabel 3. Tabel Hasil Training Regresi Logistik

Epoch	Data ke	Params	Bobot terbaru
1	1	w0	-0.0090375
1	1	w27	-0.012861
1	1	w28	-0.0137985
1	1	w29	-0.0070775
1	1	bias	-0.05
...
100	453	w29	1.0259383755821
100	453	Bias	-14.65737214692

Setelah semua tahapan pada logistic regression selesai maka didapatkan bobot dan *bias* yang sudah mencapai konvergen sebagai berikut untuk digunakan pada model.

$w_1 = 1.112295769$, $w_2 = 3.343807482$, $w_3 = 1.306493039$, $w_4 = 2.425517843$, $w_5 = 0.602807109$, $w_6 = -0.390593878$, $w_7 = 5.473738718$, $w_8 = 6.544947131$, $w_9 = -0.861022335$, $w_{10} = -3.082270344$, $w_{11} = 6.951441336$, $w_{12} = 0.615031295$, $w_{13} = 5.031788795$, $w_{14} = 4.180217621$, $w_{15} = -1.658537338$, $w_{16} = -4.270020577$, $w_{17} = -1.502080558$, $w_{18} = -0.857688098$, $w_{19} = -1.259672348$, $w_{20} = -2.527739834$, $w_{21} = 5.06263628$, $w_{22} = 5.359061658$, $w_{23} = 4.438645327$, $w_{24} = 4.750069793$, $w_{25} = 3.819948308$, $w_{26} = 0.465713857$, $w_{27} = 3.06261442$, $w_{28} = 3.922472175$, $w_{29} = 3.83756501$, $w_{30} = 1.025938376$, $b = -14.65737215$

Setelah didapatkan bobot dan bias, maka tahap selanjutnya adalah melakukan testing dengan data test yang sudah disiapkan. Model = $z = \sum_{i=0}^{29} w_i x_i + b$ dan dilanjutkan dengan fungsi aktivasi sigmoid $\hat{y} = \text{sigmoid}(z)$ Dimana setelah didapatkan nilai dari fungsi aktivasi, dilanjutkan dengan tresholding (decision treshold) $\hat{y} \in [0,1]$

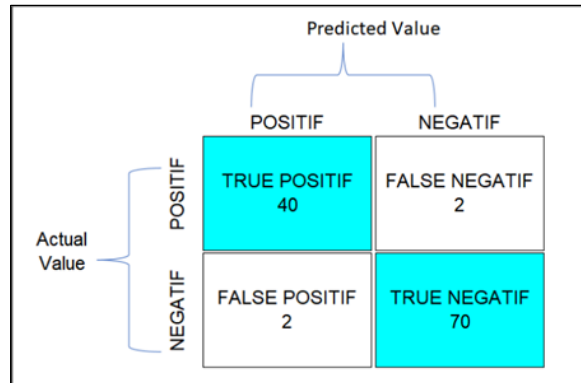
$$prediksi = \begin{cases} 1 & \text{jika } \hat{y} \geq 0.5 \\ 0 & \text{jika } \hat{y} \leq 0.5 \end{cases}$$

Tabel 4. Tabel Hasil Testing Regresi Logistik

Data ke	Fungsi aktivasi	Prediksi	Output yang diharapkan
1	$y = 0.013009860900029703 = 0.5 ? 0$	0	0
2	$y = 0.004049546888396026 = 0.5 ? 0$	0	0
3	$y = 0.7253331916837131 = 0.5 ? 1$	1	0
...
48	$y = 0.012710311884395939 = 0.5 ? 0$	0	0
49	$y = 0.006957144343695427 = 0.5 ? 0$	0	0
50	$y = 0.525078640398296 = 0.5 ? 1$	1	0
...
97	$y = 0.2373333780750071 = 0.5 ? 0$	0	1
98	$y = 0.9928861057402041 = 0.5 ? 1$	1	1
106	$y = 0.02957256414022327 = 0.5 ? 0$	0	0

107	$y = 0.1771008623116661 = 0.5 ? 0$	0	1
108	$y = 0.02492524279155417 = 0.5 ? 0$	0	0
...
115	$y = 0.6698848990882303 = 0.5 ? 1$	1	1

Evaluasi kinerja performa model Regresi Logistik diukur menggunakan *confusion matrix* untuk menghitung metrik akurasi, presisi, *recall*, dan *F1-score*.



Gambar 3. Confusion Matrix Regresi Logistik

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$= (40 + 70) / (40 + 71 + 2 + 2) = 0.9565$$

$$Precision = TP / (TP + FP)$$

$$= 40 / (40 + 2) = 0.9524$$

$$Recall = TP / (TP + FN)$$

$$= 40 / (40 + 2) = 0.9524$$

$$F1\ Score = 2 \times ((precision \times recall) / (precision + recall))$$

$$= 2 \times ((0.9524 \times 0.9524) / (0.9524 + 0.9524)) = 0.9524$$

Hasil evaluasi menunjukkan bahwa model ini mencapai nilai akurasi sebesar 95,65%.

3. HASIL DAN PEMBAHASAN

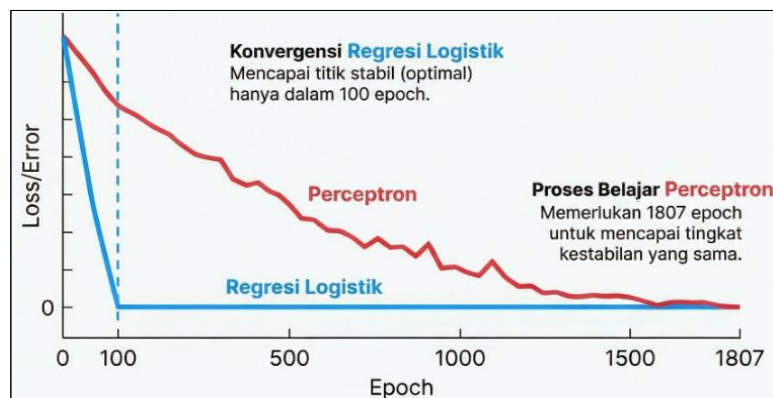
3.1 Analisis Implementasi dan Proses Pelatihan

Proses implementasi kedua model menunjukkan perbedaan signifikan dalam mencapai titik optimal. Model Perceptron klasik memerlukan waktu pelatihan yang jauh lebih lama, mencapai 1807 epoch sebelum bobot dan biasnya stabil. Hal ini mengindikasikan bahwa aturan pembaruan bobot (*delta rule*) pada Perceptron, yang hanya bergantung pada kesalahan klasifikasi biner, membutuhkan iterasi yang lebih banyak untuk menemukan garis pemisah pada dataset medis yang kompleks.

Sebaliknya, Regresi Logistik mencapai konvergensi hanya dalam 100 *epoch* dengan nilai bias akhir sebesar -14,657. Efisiensi ini didorong oleh penggunaan fungsi loss (*Log Loss*) dan algoritma *gradient descent* yang melakukan penyesuaian bobot secara kontinu berdasarkan probabilitas, bukan sekadar benar atau salah secara biner. Hal ini membuktikan bahwa pendekatan probabilistik lebih efisien secara komputasi untuk dataset *Breast Cancer Wisconsin* dibandingkan model syaraf lapis tunggal sederhana.

Tabel 5. Tabel Perbandingan Parameter dan Konvergensi Model

Parameter	Perceptron Klasik	Regresi Logistik
Fungsi Aktivasi	<i>Step Function</i> (Biner)	Sigmoid (Probabilistik)
<i>Learning Rate</i>	0,1	0,05
Inisialisasi Bobot	0	Acak (-0,01 hingga 0,01)
Jumlah <i>Epoch</i>	1907 <i>Epoch</i>	100 <i>Epoch</i>
Nilai Bias Akhir	-2,6	-14,657
Kriteria Henti	Prediksi = Lable ($y_{pred} = y$)	Titik konvergen (<i>mean loss minimal</i>)

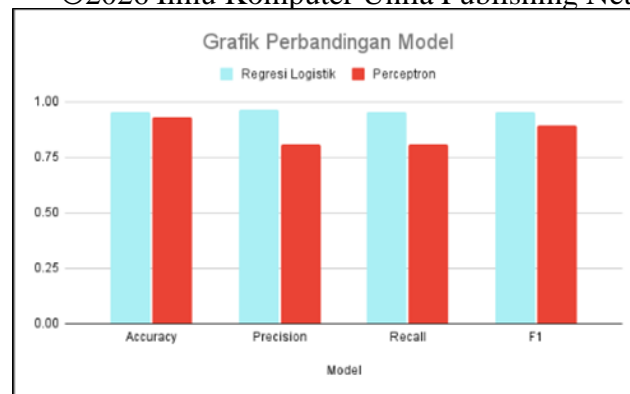


Gambar 4. Perbandingan Regresi Logistik dan Perceptron

3.2 Evaluasi Kinerja dan Perbandingan Confusion Matrik

Berdasarkan hasil pengujian pada 115 sampel data uji, performa kedua model dapat dibandingkan secara mendalam melalui metrik berikut:

- Akurasi : Regresi Logistik mengungguli Perceptron dengan akurasi 95,65% berbanding 93%. Selisih ini menunjukkan bahwa Regresi Logistik memiliki kemampuan generalisasi yang lebih baik terhadap data baru yang belum pernah dilihat sebelumnya.
- Analisis *Recall* dan Resiko Medis : perbedaan yang paling mencolok terletak pada nilai *Recall*, di mana Perceptron hanya mencapai 81%, sementara Regresi Logistik mencapai 95,24%. Secara praktis, rendahnya *recall* pada Perceptron berarti terdapat 8 pasien dengan karakter ganas (*Malignant*) yang salah terdeteksi sebagai jinak (*Benign*) atau *false negative*. Dalam konteks medis, kegagalan identifikasi ini sangat berisiko karena dapat menunda penanganan medis yang kritis bagi pasien.
- Stabilitas Prediksi : Regresi Logistik menunjukkan keseimbangan yang sangat baik antara presisi dan sensitivitas dengan F1-score sebesar 95,24%, jauh di atas perceptron yang hanya 89,4%



Gambar 5. Grafik Perbandingan Model

3.3 Analisa Karakteristik Model terhadap Data

Keunggulan Regresi Logistik berakar pada penggunaan fungsi aktivasi Sigmoid. Fungsi ini menghasilkan *output* kontinu antara 0 dan 1, yang memungkinkan model untuk menangani data yang tidak sepenuhnya terpisah secara linier. Sebagian besar fitur morfologi sel kanker memiliki tumpang tindih nilai antara kelas ganas dan jinak, sehingga fleksibilitas Sigmoid dalam memodelkan hubungan fitur numerik menjadi kunci keberhasilan klasifikasi. Di sisi lain, Perceptron klasik dibatasi oleh fungsi aktivasi step (biner) yang sangat kaku. Karena model ini hanya memiliki lapisan tunggal tanpa lapisan tersembunyi (*hidden layer*), performanya menurun secara signifikan saat menghadapi pola data medis yang kompleks dan memiliki kemiripan fitur yang tinggi antar kelas. Hasil penelitian ini menegaskan bahwa untuk klasifikasi medis awal, model linier dengan pendekatan probabilistik seperti Regresi Logistik jauh lebih stabil dan dapat diandalkan dibanding model syaraf dasar.

4. KESIMPULAN

Berdasarkan hasil analisis komparatif pada penelitian ini, dapat disimpulkan bahwa model Regresi Logistik terbukti jauh lebih efektif dan stabil dalam melakukan klasifikasi biner pada dataset *Breast Cancer Wisconsin* dibandingkan dengan model Perceptron klasik. Model Regresi Logistik menunjukkan performa yang unggul dengan pencapaian akurasi dan presisi sebesar 95,65%, serta nilai *recall* dan *F1-score* sebesar 95,24%. Sebaliknya, meskipun Perceptron klasik mencatatkan akurasi yang cukup baik sebesar 93%, model ini memiliki kelemahan signifikan pada aspek *recall* dan presisi yang hanya mencapai 81%, sehingga memiliki risiko lebih tinggi dalam kegagalan deteksi sel ganas. Keunggulan Regresi Logistik terletak pada pendekatan probabilistik melalui fungsi aktivasi sigmoid yang lebih fleksibel dalam memodelkan hubungan antar fitur numerik kompleks, sementara Perceptron klasik terbatas oleh fungsi step biner yang hanya efektif pada data yang terpisah secara linier. Oleh karena itu, Regresi Logistik lebih direkomendasikan sebagai model dasar dalam klasifikasi medis, sementara studi ini memberikan kontribusi edukatif penting mengenai keterbatasan model linier sederhana dalam menghadapi data dunia nyata yang kompleks. Untuk penelitian selanjutnya, disarankan untuk mengembangkan model yang lebih kompleks seperti Multilayer Perceptron (MLP) atau Deep Learning guna meningkatkan performa klasifikasi. Selain itu, eksplorasi teknik feature selection dan hyperparameter tuning juga dapat dilakukan untuk memperoleh hasil yang lebih optimal. Penggunaan metode cross-validation juga direkomendasikan untuk meningkatkan validitas hasil penelitian.

DAFTAR PUSTAKA

- [1] S. Arti and E. Suherlan, "E ISSN : 2809-4069 Evaluasi Kinerja Machine Learning dalam Memprediksi Kemampuan Adaptasi Mahasiswa pada Lingkungan Pembelajaran Daring," vol. 5, no. 1, pp. 50–57, 2025.
- [2] Y. M. Nimas Ratna Sari, "PENERAPAN MULTILAYER PERCEPTRON UNTUK IDENTIFIKASI KANKER PAYUDARA," *J. Cakrawala Ilm.*, vol. 2, no. 8, pp. 3261–3268, 2023.

- [3] M. H. Yuhandri and L. Mayola, "Identifikasi Pola Seleksi Penentuan Calon Wali Nagari dengan Menggunakan Artificial Neural Network Algoritma Perceptron," *J. KomtekInfo*, vol. 10, no. 4, pp. 158–164, 2023, doi: 10.35134/komtekinfo.v10i4.485.
- [4] D. Nurlina, L. S. Harahap, K. Essiva, and S. Silalahi, "Klasifikasi Lemon dan Lime Menggunakan Single Layer Perceptron Berdasarkan Ciri Fisik," vol. 1, no. 2, pp. 12–21, 2025.
- [5] L. Dwi, N. Khasanah, M. Fathurahman, and M. N. Hayati, "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," vol. 15, no. November, pp. 62–72, 2024, doi: 10.30872/eksponensial.v15i2.1301.
- [6] S. Sudriyanto, A. Fatah, M. Dafa, and W. Putra, "Evaluasi Model Machine learning untuk Prediksi Keparahan Kanker Berdasarkan Data Real-world Global," vol. 7, no. 02, pp. 922–935, 2025.
- [7] R. Meri, D. E. Putra, R. A. Efendi, and R. I. Salam, "Pengembangan Sistem Pendukung Keputusan Diagnosis Kanker Payudara Menggunakan Regresi Logistik," pp. 499–506, 2025.
- [8] S. Vanitra, B. C. Oktariadi, S. Putri, and A. Alkadri, "Penerapan Jaringan Syaraf Tiruan Backpropagation dalam Pengenalan Huruf Hijaiyah," pp. 45–52, 2025.
- [9] Z. Fatah and R. Atreji, "Klasifikasi Penyakit Gagal Jantung Menggunakan Metode Decision Tree," vol. 4, no. 3, pp. 844–854, 2025.
- [10] C. Melisa and R. Panjaitan, "Jurnal JPILKOM (Jurnal Penelitian Ilmu Komputer) Penerapan Algoritma Backpropagation dan Principal Component Analysis (PCA) pada Jaringan Syaraf Tiruan untuk Identifikasi Jenis Kulit Wajah Berdasarkan Data Kuesioner," vol. 2, no. 1, 2024.
- [11] F. Tangguh and Y. Islami, "Analisis performa algoritma Stochastic Gradient Descent (SGD) dalam mengklasifikasi tahu berformalin," vol. 3, no. 1, pp. 1–8, 2022.
- [12] W. K. Sari *et al.*, "Perbandingan Kinerja Neural Network dengan Metode Klasifikasi Tradisional dalam Mendiagnosis Penyakit Jantung: Sebuah Studi Komparatif 1," vol. 15, no. 1, pp. 3111–3117, 2023.
- [13] P. Ketinggian, A. Sungai, K. Kurniawan, and B. Ceasaro, "Perbandingan Fungsi Aktivasi Untuk Meningkatkan Kinerja Model LSTM Dalam," vol. 10, no. 1, pp. 134–143, 2024.
- [14] S. Wulandari and D. Novita, "Implementasi Feed Forward Neural Network (FFNN) dalam Memprediksi Penyakit Diabetes," vol. 3, no. 1, pp. 1–11, 2024.
- [15] P. Alkhairi, E. R. Batubara, R. Rosnelly, W. Wanayaumini, and H. S. Tambunan, "Effect of Gradient Descent With Momentum Backpropagation Training Function in Detecting Alphabet Letters," vol. 7, no. 1, pp. 574–583, 2023.