

LAPORAN TUGAS BESAR 1
IF3270 PEMBELAJARAN MESIN
Feedforward Neural Network



Disusun Oleh:

13522060 Andhita Naura Hariyanto

13522062 Salsabiila

13522082 Keanu Amadius Gonza Wrahatno

PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
2025

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR GAMBAR.....	3
DAFTAR TABEL.....	4
BAB I	
Deskripsi Persoalan.....	5
BAB II	
Pembahasan.....	7
2.1. Penjelasan Implementasi.....	7
2.1.1. Deskripsi Kelas.....	7
a. activation_function.py.....	7
b. activation_turunan.py.....	8
c. loss_function.py.....	9
e. Weight_initializer.py.....	10
f. backpropagation.py.....	11
g. ffnn.py.....	12
h. static_visualizer.py dan interactive_visualizer.py.....	13
2.1.2. Forward Propagation.....	14
2.1.3. Backward Propagation.....	14
2.2. Hasil Pengujian.....	15
2.2.1. Pengaruh Width.....	15
2.2.2. Pengaruh Depth.....	26
2.2.3. Pengaruh Fungsi Aktivasi.....	40
2.2.4. Pengaruh Learning Rate.....	48
2.2.5. Pengaruh Inisialisasi Bobot.....	59
2.2.6. Pengaruh Regularisasi.....	80
2.2.7. Perbandingan dengan Library Sklearn.....	89
2.2.8. Pengujian terhadap Dataset MNIST_784.....	93
BAB III	
Kesimpulan dan Saran.....	97
3.1. Kesimpulan.....	97
3.2. Saran.....	98
LAMPIRAN.....	99
REFERENSI.....	100

DAFTAR GAMBAR

Gambar 1.1 Visualisasi Umum FFNN

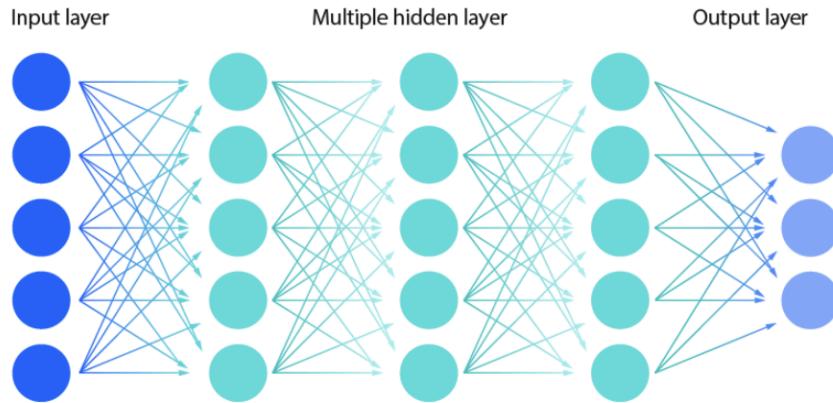
5

DAFTAR TABEL

2.2.1.1 Hidden Layer dengan Jumlah Neuron 75	16
2.2.1.2 Hidden Layer dengan Jumlah Neuron 150	18
2.2.1.3 Hidden Layer dengan Jumlah Neuron 200	21
2.2.1.4 Hidden Layer dengan Jumlah Neuron 300	23
2.2.2.1 Model dengan 2 Hidden Layer	26
2.2.2.2 Model dengan 4 Hidden Layer	29
2.2.2.3 Model dengan 6 Hidden Layer	32
2.2.2.4 Model dengan 8 Hidden Layer	36
2.2.2.1 Fungsi Aktivasi sigmoid	41
2.2.2.2 Fungsi Aktivasi tanh	43
2.2.2.3 Fungsi Aktivasi swish	46
2.2.3.1 Learning Rate 0.1	49
2.2.3.2 Learning Rate 0.001	52
2.2.3.3 Learning Rate 0.005	56
2.2.4.1. Zero Initialization	60
2.2.4.2. Random Uniform Initialization	62
2.2.4.3. Random Normal Initialization	65
2.2.4.4. Normal Xavier Initialization	68
2.2.4.5. Uniform Xavier Initialization	71
2.2.4.6. Normal He Initialization	74
2.2.4.7. Uniform He Initialization	77
2.2.6.1. Tanpa Regularisasi	81
2.2.6.2. Dengan Regularisasi L1	84
2.2.6.3. Dengan Regularisasi L2	86
2.2.6.4. Library Sklearn MLP	89
2.2.6.5. FFNN Buatan	92
2.2.6.1 dataset mnist_784	94

BAB I

Deskripsi Persoalan



Gambar 1.1 Visualisasi Umum FFNN

Pada tugas besar ini, akan diimplementasikan sebuah modul Feedforward Neural Network (FFNN). Modul ini harus dapat menerima konfigurasi jumlah neuron pada setiap *layer*, mulai dari *input layer* hingga *output layer*, sehingga struktur jaringan dapat dirancang sesuai karakteristik permasalahan yang ingin diselesaikan.

Selain itu, setiap layer dalam jaringan dapat menggunakan fungsi aktivasi yang berbeda. Fungsi aktivasi yang perlu didukung mencakup Linear, ReLU, Sigmoid, Hyperbolic Tangent (*tanh*), dan Softmax. Untuk proses pelatihan, model juga harus bisa menggunakan beberapa jenis fungsi *loss* sesuai kebutuhan, yaitu Mean Squared Error (MSE), Binary Cross-Entropy, dan Categorical Cross-Entropy.

FFNN ini juga perlu memiliki beberapa metode inisialisasi bobot dan bias, termasuk inisialisasi nol (*zero initialization*), distribusi acak uniform, dan distribusi acak normal. Setiap model yang diinisialisasi harus dapat menyimpan nilai bobot dan bias dari setiap neuron, serta menyimpan nilai gradiennya untuk keperluan proses pelatihan.

Agar model ini dapat dianalisis dengan baik, perlu disediakan fitur visualisasi yang menampilkan struktur jaringan lengkap dengan bobot dan gradien masing-masing neuron dalam bentuk grafis. Selain itu, model juga harus memiliki fungsi untuk menampilkan distribusi bobot dan distribusi gradien dari tiap layer, sehingga memudahkan dalam memahami dinamika pelatihan. Model juga harus memiliki kemampuan untuk disimpan (*save*) dan dimuat kembali (*load*) agar dapat digunakan ulang tanpa pelatihan dari awal.

Secara fungsional, model ini wajib mendukung proses forward propagation untuk data dalam bentuk batch, serta backward propagation untuk menghitung gradien berdasarkan data batch tersebut. Setelah gradien dihitung, bobot diperbarui menggunakan metode gradient descent. Untuk melatih model, pengguna harus bisa mengatur beberapa parameter penting seperti *batch size*, *learning rate*, jumlah epoch, dan opsi *verbose* untuk menampilkan progress pelatihan. Proses pelatihan ini akan menghasilkan histori berupa *training loss* dan *validation loss* pada setiap epoch, yang dapat digunakan untuk mengevaluasi performa model selama proses pelatihan.

BAB II

Pembahasan

2.1. Penjelasan Implementasi

2.1.1. Deskripsi Kelas

Berikut merupakan kelas pendukung yang digunakan dalam implementasi Feedforward Neural Network (FFNN).

a. activation_function.py

Kelas ActivationFunction berisi kumpulan metode statis yang digunakan untuk menerapkan berbagai fungsi aktivasi. Berikut adalah daftar fungsi aktivasi yang tersedia dalam kelas ini:

- Linear
Fungsi ini mengembalikan nilai input apa adanya, tanpa perubahan apa pun. Karena sifatnya yang sederhana, fungsi ini kurang cocok untuk digunakan di hidden layer dalam jaringan yang kompleks.
- ReLU (Rectified Linear Unit)
Fungsi ini mengembalikan nilai input jika positif, dan nol jika negatif. Fungsi umumnya dipilih untuk hidden layer karena membantu mengatasi masalah vanishing gradient dan mempercepat konvergensi.
- Sigmoid
Fungsi ini mengubah nilai input menjadi rentang antara 0 dan 1 dan biasanya digunakan di output layer untuk klasifikasi biner karena menghasilkan output dalam bentuk probabilitas.
- Hyperbolic Tangent (tanh)
Fungsi ini merupakan versi terpusat dari sigmoid yang menghasilkan output di antara -1 dan 1.
- Softmax
Fungsi ini mengubah vektor input menjadi distribusi probabilitas, sehingga jumlah seluruh output menjadi 1 dan biasanya digunakan di output layer untuk klasifikasi multi-kelas, di mana setiap neuron mewakili satu kelas.
- Softplus
Fungsi aktivasi yang merupakan versi halus dari ReLU, yang didefinisikan sebagai $\log(1 + \exp(x))$. Fungsi ini selalu menghasilkan output positif dan bersifat diferensiabel di seluruh domain sehingga lebih stabil untuk pelatihan namun lebih mahal secara komputasi dibanding ReLU.
- ELU (Exponential Linear Unit)

Fungsi ini memiliki kelebihan ReLU dan mengatasi kekurangannya dengan menghasilkan nilai negatif halus ketika input kurang dari nol. Hal ini membantu menjaga nilai aktivasi tetap mendekati nol yang dapat mempercepat pelatihan.

- SELU (Scaled Exponential Linear Unit)

Merupakan versi ELU yang diskalakan secara khusus. SELU digunakan dalam arsitektur jaringan self-normalizing karena memiliki sifat mempertahankan statistik aktivasi (mean dan varians) yang stabil selama pelatihan.

- PReLU (Parametric Rectified Linear Unit)

Variasi dari ReLU di mana kemiringan pada bagian negatif dari fungsi bukan nol, tetapi merupakan parameter masukan α .

- Swish

Fungsi aktivasi non-monotonik yang didefinisikan sebagai $x \cdot \text{sigmoid}(x)$. Swish memiliki performa lebih baik daripada ReLU dalam beberapa arsitektur *deep learning* karena menghasilkan aktivasi yang lebih halus dan fleksibel.

b. activation_turunan.py

Kelas ActivationDerivative berisi implementasi turunan dari berbagai fungsi aktivasi yang tersedia pada kelas ActivationFunction. Turunan-turunan ini digunakan dalam proses backpropagation untuk menghitung gradien dari fungsi aktivasi terhadap input-nya, yang kemudian digunakan dalam pembaruan bobot selama pelatihan model. Kelas ini terdiri dari beberapa metode statis, yaitu:

- Linear

Mengembalikan nilai 1 untuk seluruh input, karena turunan dari fungsi linear $f(x) = x$ adalah 1.

- ReLU (Rectified Linear Unit)

Mengembalikan 1 jika input lebih besar dari nol, dan 0 jika input kurang dari atau sama dengan nol. Turunan ini membantu menyaring gradien selama pelatihan.

- Sigmoid

Menghitung turunan sigmoid berdasarkan output-nya, yaitu $\sigma(x)(1 - \sigma(x))$. Nilai ini merepresentasikan sensitivitas perubahan output terhadap input.

- Hyperbolic Tangent (\tanh)

Mengembalikan nilai turunan dari fungsi \tanh terhadap input, yaitu $1 - \tanh^2(x)$.

- Softmax

Untuk keperluan backpropagation, turunan softmax biasanya dikombinasikan langsung dengan turunan dari fungsi loss (seperti categorical cross-entropy) karena hasilnya berupa matriks Jacobian. Oleh karena itu, implementasinya secara eksplisit jarang dilakukan kecuali dalam bentuk gabungan.

- Softplus

Menghitung turunan dari softplus sebagai $\frac{1}{1 + \exp(-x)}$ yang identik dengan sigmoid.

- ELU (Exponential Linear Unit)
Turunannya adalah 1 untuk input positif, dan $\alpha \cdot \exp(x)$ untuk input negatif.
- SELU (Scaled Exponential Linear Unit)
Sama seperti ELU, namun dikalikan dengan faktor skala tertentu. Turunannya disesuaikan dengan skala dan α pada fungsi SELU.
- PReLU (Parametric Rectified Linear Unit)
Mengembalikan 1 untuk input positif, dan nilai parameter α untuk input negatif.
- Swish
Menggunakan bentuk turunan gabungan dari fungsi $x \cdot \text{sigmoid}(x)$, yaitu $\sigma(x) + x \cdot \sigma(x)(1 - \sigma(x))$.

c. loss_function.py

Kelas LossFunction berisi kumpulan metode statis yang digunakan untuk menghitung loss functions. Kelas ini menyediakan tiga jenis algoritma loss function:

- Mean Squared Error (MSE)
Digunakan untuk regresi dengan menghitung rata-rata dari jumlah kuadrat selisih antara nilai prediksi (y_{pred}) dan nilai sebenarnya (y_{true}).
- Binary Cross-Entropy (BCE)
Digunakan untuk klasifikasi biner dengan menghitung rata-rata negatif log likelihood antara prediksi probabilitas dan label sebenarnya. Nilai prediksi dikonversi dengan clipping untuk menghindari logaritma dari nol.
- Categorical Cross-Entropy (CCE)
Digunakan untuk klasifikasi multi kelas dengan menghitung rata-rata negatif log likelihood dari distribusi probabilitas prediksi terhadap label sebenarnya yang direpresentasikan sebagai one hot encoding. Nilai prediksi dikonversi dengan clipping untuk menghindari logaritma dari nol.
- L1 Regularization
Digunakan untuk mengkalkulasikan penambahan penalti untuk metode regularisasi L1, yaitu $\lambda \sum_{i=1}^N |\theta_i|$.
- L2 Regularization
Digunakan untuk mengkalkulasikan penambahan penalti untuk metode regularisasi L2, yaitu $\lambda \sum_{i=1}^N \theta_i^2$.

d. loss_turunan.py

Kelas LossDerivative berisi implementasi turunan dari berbagai loss functions yang ada pada kelas LossFunction. Turunan ini berguna untuk menghitung gradien pada backpropagation. Kelas ini terdiri dari beberapa metode statis, yaitu:

- Mean Squared Error (MSE)
Menghitung turunan fungsi MSE terhadap y_{pred}
- Binary Cross-Entropy (BCE)
Menghitung turunan fungsi BCE terhadap y_{pred} .
- Categorical Cross-Entropy (CCE)
Menghitung turunan fungsi CCE terhadap y_{pred} , label sebenarnya direpresentasikan sebagai one-hot encoding.
- L1 Regularization
Menghitung turunan untuk kalkulasi penalti regularisasi L1.
- L2 Regularization
Menghitung turunan untuk kalkulasi penalti regularisasi L2.

e. Weight_initializer.py

Kelas WeightInitializer berisi implementasi metode penginisialisasi seluruh bobot input dan bias *neural network* yang terbentuk. Terdapat lima metode inisialisasi bobot yang kelas WeightInitializer sediakan :

- Zero Initializer
Menginisialisasi bobot input dan bobot bias dengan nilai nol.
- Random Uniform Initializer
Menginisialisasi bobot secara acak dengan batas bawah rentang nilai ditentukan parameter *lower_bound*, batas atas ditentukan parameter *upper_bound*, *reproducibility* yang ditentukan parameter *seed*, dan menggunakan distribusi seragam untuk rentang nilai yang ditentukan.
- Random Normal Initializer
Menginisialisasi bobot secara acak menggunakan distribusi normal dengan rata-rata nilai ditentukan parameter *mean*, varians nilai ditentukan parameter *variance*, serta *reproducibility* yang ditentukan parameter *seed*.
- Xavier Initializer
Menginisialisasi bobot secara acak dengan rentang nilai

$$(-\sqrt{\frac{6}{input\ nodes + output\ nodes}}, \sqrt{\frac{6}{input\ nodes + output\ nodes}})$$

Pembatasan rentang pada Xavier Initialization menjaga varian bobot tetap terkontrol karena rentang yang dihasilkan tetap dapat memberikan nilai batas distribusi yang terjaga. Xavier Initialization mencegah terjadinya nilai gradien yang terlalu besar (exploding gradient) jika jumlah neuron membesar atau bobot

diinisialisasi dengan nilai yang terlalu besar dan mencegah terjadinya nilai gradien yang terlalu kecil (*vanishing gradient*) akibat nilai hasil fungsi aktivasi yang mengecil. Pengontrolan nilai varian oleh Xavier Initialization membuat gradien tetap stabil saat melakukan backpropagation. Secara umum, metode inisialisasi bobot dengan Xavier Initialization cocok digunakan untuk neural network dengan fungsi aktivasi sigmoid dan tanh, namun kurang optimal untuk ReLU karena Xavier Initialization mengasumsikan bahwa fungsi aktivasi memiliki distribusi simetris di sekitar nol seperti sigmoid dan tanh.

- He Initializer

Menginisialisasi bobot secara acak menggunakan distribusi gaussian normal dengan standar deviasi rentang nilai yang ditentukan oleh

$$\text{Standar Deviasi} = \sqrt{\frac{2}{\text{input nodes}}}$$

He Initialization mencegah terjadinya nilai gradien yang terlalu besar (*exploding gradient*) jika jumlah neuron membesar atau bobot diinisialisasi dengan nilai yang terlalu besar dan mencegah terjadinya nilai gradien yang terlalu kecil (*vanishing gradient*) akibat nilai hasil fungsi aktivasi yang mengecil. Pengontrolan nilai varian oleh He Initialization membuat gradien tetap stabil saat melakukan backpropagation. Secara umum, metode inisialisasi bobot dengan He Initialization cocok digunakan untuk *neural network* dengan fungsi aktivasi ReLU, namun kurang optimal untuk sigmoid dan tanh karena tidak mempertimbangkan distribusi simetris di sekitar nol.

f. backpropagation.py

Penjelasan

Kelas ini mengimplementasikan algoritma backpropagation untuk menghitung gradien dan memperbarui bobot serta bias dalam model. Ada beberapa method didalamnya:

- Compute Gradients

Algoritma untuk menghitung gradien dari bobot dan bias menggunakan algoritma backpropagation. Dilakukan dengan menginisialisasi variabel untuk menyimpan gradien. Lalu menghitung loss sesuai metode yang digunakan.

Setelah itu, dilakukan proses iterasi mundur dari lapisan output ke lapisan awal. Gradien bobot dihitung sebagai hasil perkalian antara aktivasi lapisan sebelumnya dengan delta. Gradien bias dihitung dengan menjumlahkan nilai delta. Delta untuk lapisan sebelumnya diperbarui menggunakan bobot lapisan saat ini.

- Update Weight

Algoritma untuk memperbarui bobot dan bias berdasarkan gradien yang dihitung pada compute gradients. Bobot diperbarui dengan dikurangi dari hasil

perkalian learning rate dengan gradien bobot. Bias diperbarui dengan dikurangi dari hasil perkalian learning rate dengan gradien bias.

- Backward

Metode ini sebagai main proses backpropagation dengan memanggil dua metode sebelumnya. Pertama memanggil `compute_gradients` untuk menghitung gradien dari bobot dan bias. Kedua memanggil `update_weights` untuk memperbarui bobot dan bias menggunakan gradien yang telah dihitung lalu mengembalikan nilai gradien sebagai hasil akhir.

g. `ffnn.py`

Kelas FFNN merupakan kelas utama yang memiliki

- Layer Sizes

Menyimpan jumlah neuron di setiap lapisan jaringan.

- Activation Functions

Menyimpan fungsi aktivasi yang digunakan di setiap lapisan.

- Loss Function

Metode inisialisasi bobot (zero, uniform, normal, xavier, atau he)

- Weight Init

Fungsi loss yang digunakan (mse, bce, atau cce)

- Seed

Nilai seed untuk randomization dalam inisialisasi bobot

- Weights

Menyimpan bobot antar layer

- Biases

Menyimpan bias antar layer

- Weight Gradients

Menyimpan gradien dari bobot.

- Bias Gradients

Menyimpan gradien dari bias

- Pre-Activations

Menyimpan nilai sebelum aktivasi di setiap lapisan

- Activations

Menyimpan nilai setelah aktivasi di setiap lapisan

- History

Menyimpan riwayat loss selama proses training

Penjelasan

- Setup Weights

Menginisialisasi bobot dan bias berdasarkan metode inisialisasi yang dipilih

- Forward

Melakukan perhitungan forward dengan menghitung nilai aktivasi di setiap lapisan dengan menerapkan bobot, bias, dan fungsi aktivasi.

- Backward
Melakukan backward pass untuk menghitung gradien loss terhadap bobot dan bias dengan menggunakan BackPropagation untuk mengupdate bobot dan bias berdasarkan gradien.
- Train
Melatih jaringan saraf dengan menggunakan data training dan menyimpan nilai loss selama proses training.
- Predict
Menggunakan forward pass untuk memprediksi output berdasarkan input X
- Print Model
Mencetak struktur jaringan saraf, termasuk jumlah neuron, fungsi aktivasi, bobot, dan bias. Sebagai debugging
- Save
Menyimpan model ke dalam file JSON, mencakup informasi seperti ukuran lapisan, fungsi aktivasi, bobot, dan bias.
- Load
Load model dari file JSON, mencakup informasi seperti ukuran lapisan, fungsi aktivasi, bobot, dan bias.
- Visualize Model
Menggunakan InteractiveVisualizer untuk menampilkan struktur jaringan saraf.
- Visualize Weight Distribution
Menggunakan InteractiveVisualizer untuk menampilkan distribusi bobot dalam jaringan saraf.
- Visualize Gradient Weight Distribution
Menggunakan InteractiveVisualizer untuk menampilkan distribusi gradien bobot.
- Visualize Loss Curve
Menggunakan InteractiveVisualizer untuk menampilkan grafik loss selama proses training.

h. static_visualizer.py dan interactive_visualizer.py

Kelas Visualizer berisi implementasi metode visualisasi *neural network* yang dibentuk beserta visualisasi distribusi bobot beserta gradien bobot. Terdapat tiga fungsi visualisasi pada kelas Visualizer :

- Network Visualizer
Melakukan visualisasi terhadap arsitektur neural network yang dihasilkan. Komponen visualisasi terdiri atas neuron dari seluruh layer dengan labelnya dan

disertai juga informasi bobot tiap node, bobot tiap bias, gradien bobot tiap node, dan gradien bobot tiap bias.

- Plot Weight Distribution

Melakukan visualisasi terhadap distribusi bobot node dan bias dari tiap layer menggunakan bar histogram serta *kernel density estimator line* untuk menunjukkan distribusi probabilitasnya. Pada visualisasi, terdapat juga informasi statistik mean, nilai minimum bobot, nilai maksimum bobot, nilai standar deviasi bobot, serta nilai dan bentuk *skewness* dari bobot seluruh node dan bias di layer terkait.

- Plot Gradient Weight Distribution

Melakukan visualisasi terhadap distribusi gradient bobot node dan bias dari tiap layer menggunakan bar histogram serta *kernel density estimator line* untuk menunjukkan distribusi probabilitasnya. Pada visualisasi, terdapat juga informasi statistik mean, nilai minimum gradien bobot, nilai maksimum gradien bobot, nilai standar deviasi gradien bobot, serta nilai dan bentuk *skewness* dari gradien bobot seluruh node dan bias di layer terkait.

- Plot Loss Curve

Melakukan visualisasi terhadap nilai loss yang dimiliki model selama proses training dalam bentuk grafik. Grafik menampilkan nilai loss yang didapat pada setiap epoch.

2.1.2. Forward Propagation

Dalam *Feedforward Neural Network* (FFNN), terdapat *input layer*, *hidden layer*, serta *output layer*. Pada fase *forward propagation*, data *input* akan melewati satu atau lebih *hidden layer* hingga mencapai ke *output layer*. Di setiap *layer*, akan dihitung jumlah hasil kali bobot (*weight*) dengan data *input* dan bias. Jumlah bobot tersebut kemudian akan dimasukkan ke dalam fungsi aktivasi *layer* untuk menghasilkan keluaran akhir untuk *layer* tersebut. Proses perhitungan pada *layer* dilakukan secara berulang dengan menggunakan fungsi aktivasi yang sama maupun berbeda hingga mendapatkan keluaran akhir dari *output layer*..

2.1.3. Backward Propagation

Setelah proses forward propagation selesai, dilakukan proses *backward propagation* dari *layer output* ke *input layer*. *Backward propagation* dilakukan untuk belajar dari kesalahan sehingga memperkecil tingkat error ketika memproses data. Jika terdapat perbedaan antara output yang dihasilkan dan output yang seharusnya, backpropagation akan menghitung error tersebut. Lalu akan dilakukan update weight untuk menyesuaikan bobot dalam jaringan agar error yang dihasilkan juga berkurang. Jika error berkurang maka tingkat prediksi dari model akan meningkat. Proses forward dan backward ini dilakukan secara

berulang sehingga model terus belajar dan memperbaiki bobot sehingga hasil prediksi semakin akurat.

Langkah-langkah yang dilakukan adalah menghitung gradien. Gradien bobot didapat dari hasil perkalian antara aktivasi sebelumnya dengan delta saat ini. Delta didapat dengan menghitung turunan dari fungsi loss terhadap output dan turunan fungsi aktivasi. Gradien bias juga dihitung dengan menjumlahkan nilai dari delta. Lalu delta untuk layer sebelumnya dihitung lagi dengan mengalikan delta saat ini dengan bobot layer saat ini, lalu dikalikan turunan fungsi aktivasi.

Lalu bobot diperbarui dengan menggunakan gradien bobot dikali dengan learning rate. Dimana hasil perkalian ini akan dikurangkan pada bobot saat ini. Bias juga diperbarui dengan rumus yang sama.

2.2. Hasil Pengujian

2.2.1. Pengaruh *Width*

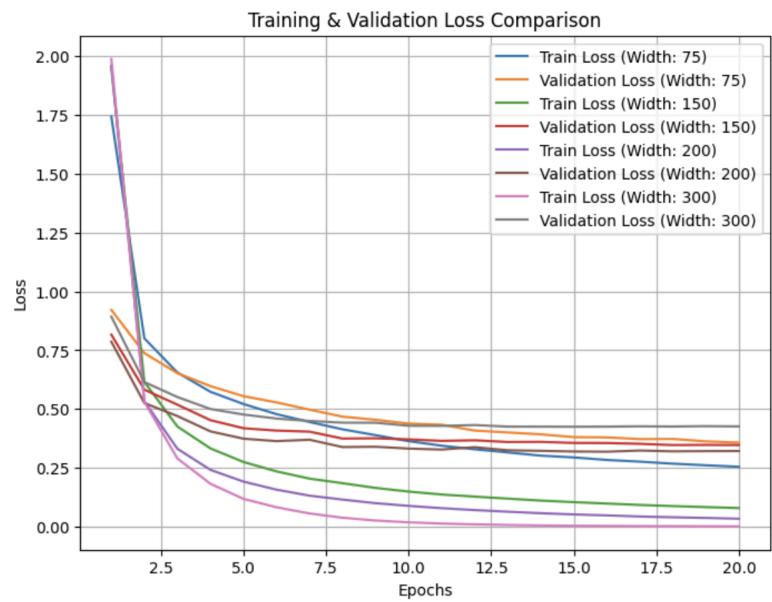
Parameter kontrol yang digunakan untuk membandingkan pengaruh antara metode inisialisasi bobot yang diimplementasikan adalah :

Jumlah layer	:	4 (input layer, hidden layer 1, hidden layer 2, output layer)
Inisialisasi Bobot	:	Normal He Initialization
Fungsi Aktivasi	:	ReLU
Fungsi loss	:	Categorical Cross-Entropy
Batch size	:	64
Learning rate	:	0.001
Epoch maksimal	:	50
Fungsi Aktivasi untuk layer terakhir	untuk	softmax

2.2.1.1 Hidden Layer dengan Jumlah Neuron 75

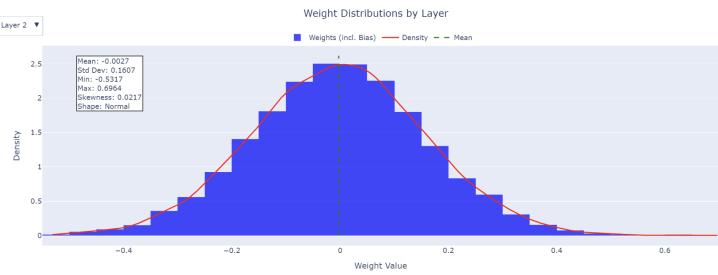
Hasil akhir prediksi		Confusion Matrix																																																																																		
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9	0	- 1400																																																																							
		1341	0	5	1	1	15	7	3	29	0	- 1200																																																																								
		1 - 0	1509	8	8	0	1	1	1	23	5	- 1000																																																																								
		2 - 6	7	1273	5	8	3	6	17	38	5	- 800																																																																								
		3 - 2	1	33	1171	0	32	2	11	70	17	- 600																																																																								
		4 - 3	1	12	0	1267	1	28	7	18	58	- 400																																																																								
		5 - 22	2	7	48	2	1092	17	5	51	15	- 200																																																																								
		6 - 15	3	12	1	7	10	1346	1	22	1	- 0																																																																								
		7 - 6	6	28	5	14	6	2	1333	16	47																																																																									
		8 - 5	3	15	24	5	14	10	4	1258	24																																																																									
		9 - 7	4	0	27	63	8	2	21	25	1279																																																																									
		0	1	2	3	4	5	6	7	8	9																																																																									
Skor validasi		Weighted F1-Score: 0.9195 Macro F1-Score: 0.9184 Cohen's Kappa: 0.9102 ROC-AUC (Weighted): 0.9916 Validation Log Loss: 0.3584 Classification Report: <table> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr><td>0</td><td>0.95</td><td>0.96</td><td>0.95</td><td>1402</td></tr> <tr><td>1</td><td>0.98</td><td>0.97</td><td>0.98</td><td>1556</td></tr> <tr><td>2</td><td>0.91</td><td>0.93</td><td>0.92</td><td>1368</td></tr> <tr><td>3</td><td>0.91</td><td>0.87</td><td>0.89</td><td>1339</td></tr> <tr><td>4</td><td>0.93</td><td>0.91</td><td>0.92</td><td>1395</td></tr> <tr><td>5</td><td>0.92</td><td>0.87</td><td>0.89</td><td>1261</td></tr> <tr><td>6</td><td>0.95</td><td>0.95</td><td>0.95</td><td>1418</td></tr> <tr><td>7</td><td>0.95</td><td>0.91</td><td>0.93</td><td>1463</td></tr> <tr><td>8</td><td>0.81</td><td>0.92</td><td>0.86</td><td>1362</td></tr> <tr><td>9</td><td>0.88</td><td>0.89</td><td>0.89</td><td>1436</td></tr> <tr><td>accuracy</td><td></td><td></td><td></td><td>0.92</td></tr> <tr><td>macro avg</td><td></td><td>0.92</td><td>0.92</td><td>0.92</td></tr> <tr><td>weighted avg</td><td></td><td>0.92</td><td>0.92</td><td>0.92</td></tr> </tbody> </table>														precision	recall	f1-score	support	0	0.95	0.96	0.95	1402	1	0.98	0.97	0.98	1556	2	0.91	0.93	0.92	1368	3	0.91	0.87	0.89	1339	4	0.93	0.91	0.92	1395	5	0.92	0.87	0.89	1261	6	0.95	0.95	0.95	1418	7	0.95	0.91	0.93	1463	8	0.81	0.92	0.86	1362	9	0.88	0.89	0.89	1436	accuracy				0.92	macro avg		0.92	0.92	0.92	weighted avg		0.92	0.92	0.92
	precision	recall	f1-score	support																																																																																
0	0.95	0.96	0.95	1402																																																																																
1	0.98	0.97	0.98	1556																																																																																
2	0.91	0.93	0.92	1368																																																																																
3	0.91	0.87	0.89	1339																																																																																
4	0.93	0.91	0.92	1395																																																																																
5	0.92	0.87	0.89	1261																																																																																
6	0.95	0.95	0.95	1418																																																																																
7	0.95	0.91	0.93	1463																																																																																
8	0.81	0.92	0.86	1362																																																																																
9	0.88	0.89	0.89	1436																																																																																
accuracy				0.92																																																																																
macro avg		0.92	0.92	0.92																																																																																
weighted avg		0.92	0.92	0.92																																																																																

Grafik loss

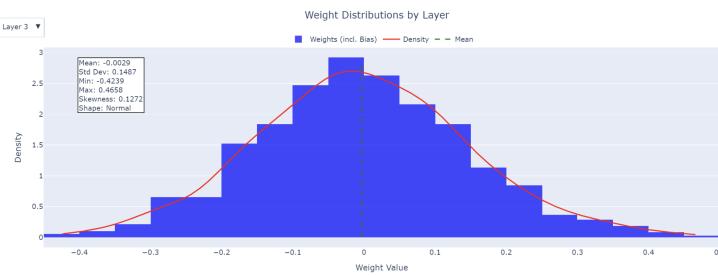


Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

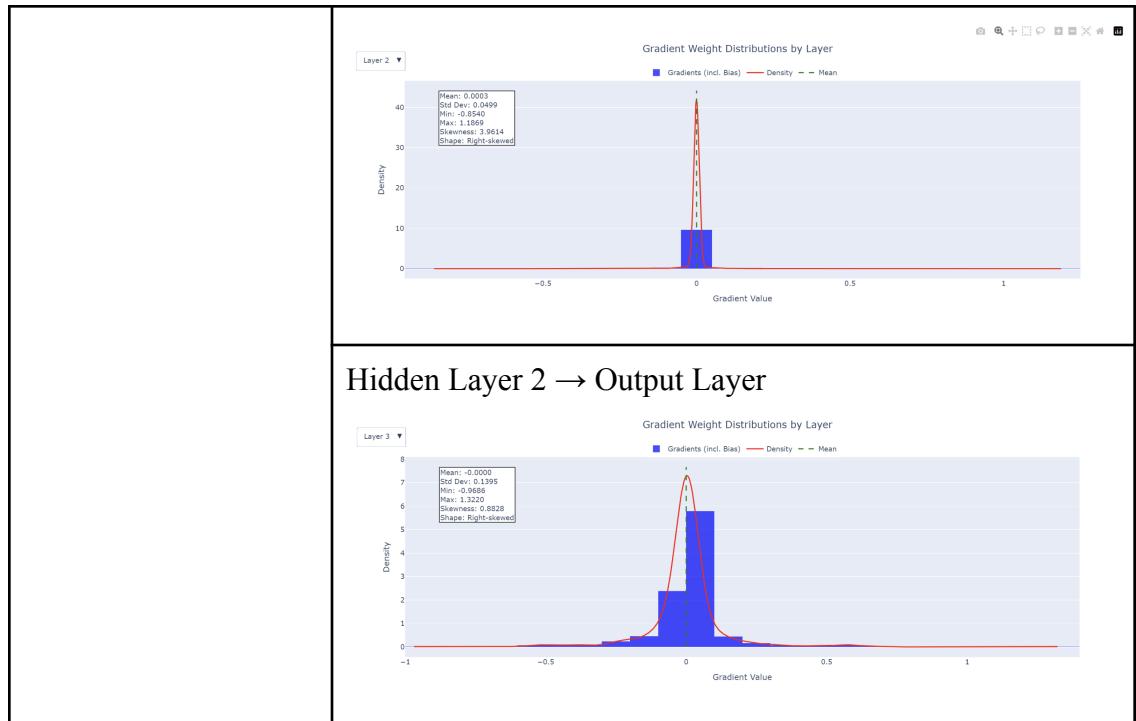


Hidden Layer 2 → Output Layer

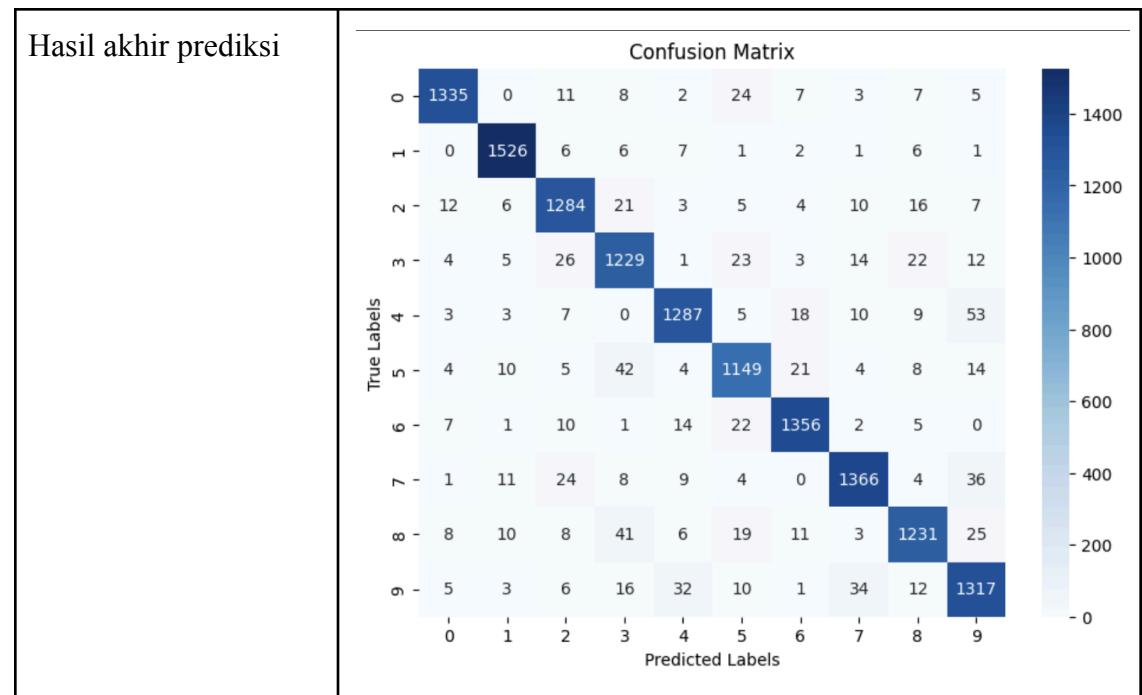


Distribusi gradien bobot akhir

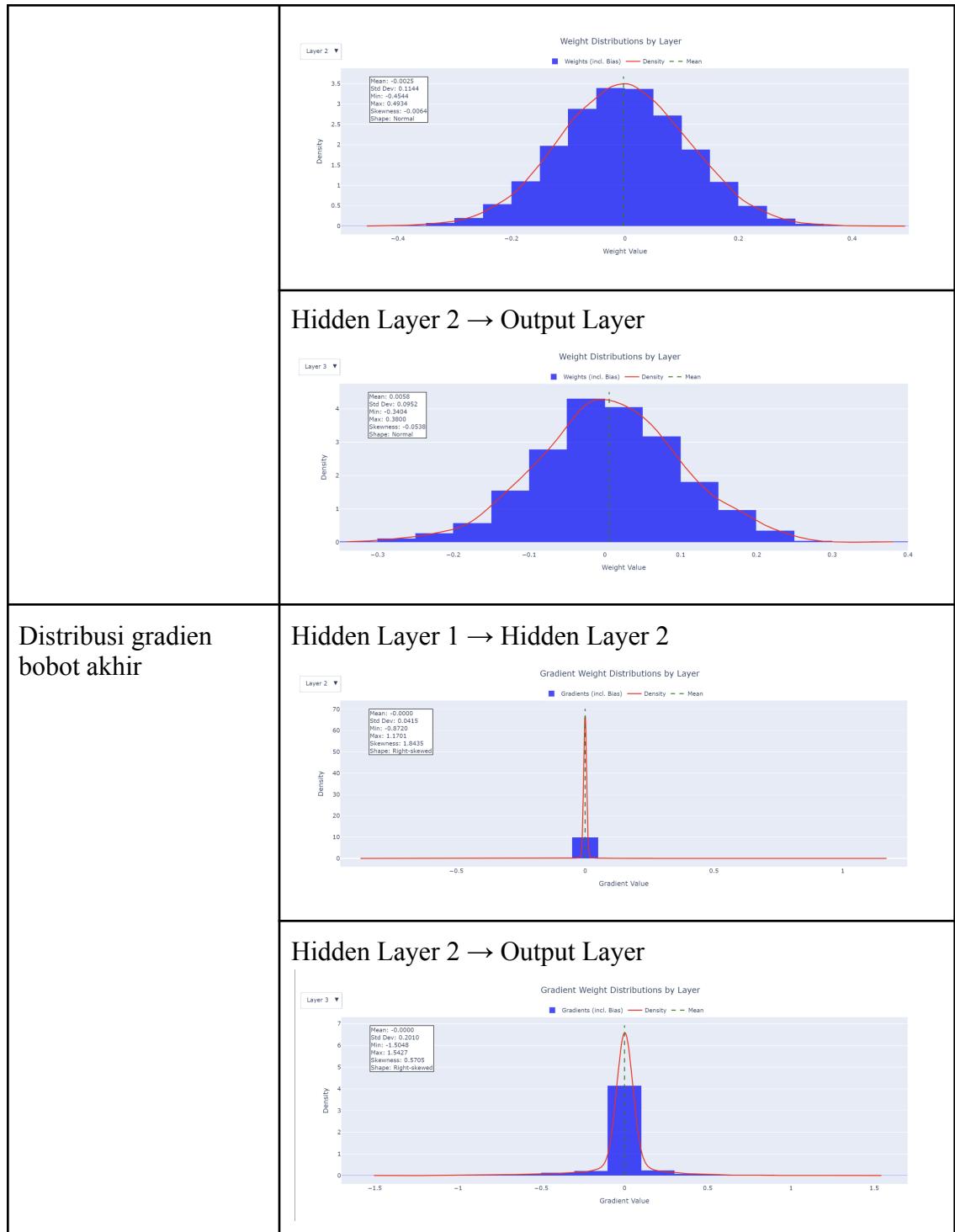
Hidden Layer 1 → Hidden Layer 2



2.2.1.2 Hidden Layer dengan Jumlah Neuron 150



Skor validasi	<p>Weighted F1-Score: 0.9343 Macro F1-Score: 0.9335 Cohen's Kappa: 0.9270 ROC-AUC (Weighted): 0.9951</p> <p>Validation Log Loss: 0.3482</p> <table border="1"> <thead> <tr> <th colspan="5">Classification Report:</th> </tr> <tr> <th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr> </thead> <tbody> <tr><td>0</td><td>0.97</td><td>0.95</td><td>0.96</td><td>1402</td></tr> <tr><td>1</td><td>0.97</td><td>0.98</td><td>0.97</td><td>1556</td></tr> <tr><td>2</td><td>0.93</td><td>0.94</td><td>0.93</td><td>1368</td></tr> <tr><td>3</td><td>0.90</td><td>0.92</td><td>0.91</td><td>1339</td></tr> <tr><td>4</td><td>0.94</td><td>0.92</td><td>0.93</td><td>1395</td></tr> <tr><td>5</td><td>0.91</td><td>0.91</td><td>0.91</td><td>1261</td></tr> <tr><td>6</td><td>0.95</td><td>0.96</td><td>0.95</td><td>1418</td></tr> <tr><td>7</td><td>0.94</td><td>0.93</td><td>0.94</td><td>1463</td></tr> <tr><td>8</td><td>0.93</td><td>0.90</td><td>0.92</td><td>1362</td></tr> <tr><td>9</td><td>0.90</td><td>0.92</td><td>0.91</td><td>1436</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.93</td><td>14000</td></tr> <tr><td>macro avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>14000</td></tr> <tr><td>weighted avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>14000</td></tr> </tbody> </table>	Classification Report:						precision	recall	f1-score	support	0	0.97	0.95	0.96	1402	1	0.97	0.98	0.97	1556	2	0.93	0.94	0.93	1368	3	0.90	0.92	0.91	1339	4	0.94	0.92	0.93	1395	5	0.91	0.91	0.91	1261	6	0.95	0.96	0.95	1418	7	0.94	0.93	0.94	1463	8	0.93	0.90	0.92	1362	9	0.90	0.92	0.91	1436	accuracy			0.93	14000	macro avg	0.93	0.93	0.93	14000	weighted avg	0.93	0.93	0.93	14000															
Classification Report:																																																																																											
	precision	recall	f1-score	support																																																																																							
0	0.97	0.95	0.96	1402																																																																																							
1	0.97	0.98	0.97	1556																																																																																							
2	0.93	0.94	0.93	1368																																																																																							
3	0.90	0.92	0.91	1339																																																																																							
4	0.94	0.92	0.93	1395																																																																																							
5	0.91	0.91	0.91	1261																																																																																							
6	0.95	0.96	0.95	1418																																																																																							
7	0.94	0.93	0.94	1463																																																																																							
8	0.93	0.90	0.92	1362																																																																																							
9	0.90	0.92	0.91	1436																																																																																							
accuracy			0.93	14000																																																																																							
macro avg	0.93	0.93	0.93	14000																																																																																							
weighted avg	0.93	0.93	0.93	14000																																																																																							
Grafik loss	<p>Training & Validation Loss Comparison</p> <table border="1"> <caption>Data for Training & Validation Loss Comparison</caption> <thead> <tr> <th>Epoch</th> <th>75 Width Train Loss</th> <th>75 Width Validation Loss</th> <th>150 Width Train Loss</th> <th>150 Width Validation Loss</th> <th>200 Width Train Loss</th> <th>200 Width Validation Loss</th> <th>300 Width Train Loss</th> <th>300 Width Validation Loss</th> </tr> </thead> <tbody> <tr><td>0</td><td>2.00</td><td>0.90</td><td>0.90</td><td>0.90</td><td>0.90</td><td>0.90</td><td>0.90</td><td>0.90</td></tr> <tr><td>2.5</td><td>0.80</td><td>0.70</td><td>0.65</td><td>0.60</td><td>0.60</td><td>0.60</td><td>0.55</td><td>0.50</td></tr> <tr><td>5.0</td><td>0.60</td><td>0.50</td><td>0.45</td><td>0.40</td><td>0.40</td><td>0.40</td><td>0.35</td><td>0.30</td></tr> <tr><td>7.5</td><td>0.50</td><td>0.40</td><td>0.35</td><td>0.30</td><td>0.30</td><td>0.30</td><td>0.25</td><td>0.20</td></tr> <tr><td>10.0</td><td>0.40</td><td>0.35</td><td>0.30</td><td>0.25</td><td>0.25</td><td>0.25</td><td>0.20</td><td>0.15</td></tr> <tr><td>12.5</td><td>0.35</td><td>0.30</td><td>0.25</td><td>0.20</td><td>0.20</td><td>0.20</td><td>0.15</td><td>0.10</td></tr> <tr><td>15.0</td><td>0.30</td><td>0.25</td><td>0.20</td><td>0.15</td><td>0.15</td><td>0.15</td><td>0.10</td><td>0.05</td></tr> <tr><td>17.5</td><td>0.28</td><td>0.24</td><td>0.20</td><td>0.14</td><td>0.14</td><td>0.14</td><td>0.08</td><td>0.04</td></tr> <tr><td>20.0</td><td>0.25</td><td>0.22</td><td>0.18</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.05</td><td>0.02</td></tr> </tbody> </table>	Epoch	75 Width Train Loss	75 Width Validation Loss	150 Width Train Loss	150 Width Validation Loss	200 Width Train Loss	200 Width Validation Loss	300 Width Train Loss	300 Width Validation Loss	0	2.00	0.90	0.90	0.90	0.90	0.90	0.90	0.90	2.5	0.80	0.70	0.65	0.60	0.60	0.60	0.55	0.50	5.0	0.60	0.50	0.45	0.40	0.40	0.40	0.35	0.30	7.5	0.50	0.40	0.35	0.30	0.30	0.30	0.25	0.20	10.0	0.40	0.35	0.30	0.25	0.25	0.25	0.20	0.15	12.5	0.35	0.30	0.25	0.20	0.20	0.20	0.15	0.10	15.0	0.30	0.25	0.20	0.15	0.15	0.15	0.10	0.05	17.5	0.28	0.24	0.20	0.14	0.14	0.14	0.08	0.04	20.0	0.25	0.22	0.18	0.12	0.12	0.12	0.05	0.02
Epoch	75 Width Train Loss	75 Width Validation Loss	150 Width Train Loss	150 Width Validation Loss	200 Width Train Loss	200 Width Validation Loss	300 Width Train Loss	300 Width Validation Loss																																																																																			
0	2.00	0.90	0.90	0.90	0.90	0.90	0.90	0.90																																																																																			
2.5	0.80	0.70	0.65	0.60	0.60	0.60	0.55	0.50																																																																																			
5.0	0.60	0.50	0.45	0.40	0.40	0.40	0.35	0.30																																																																																			
7.5	0.50	0.40	0.35	0.30	0.30	0.30	0.25	0.20																																																																																			
10.0	0.40	0.35	0.30	0.25	0.25	0.25	0.20	0.15																																																																																			
12.5	0.35	0.30	0.25	0.20	0.20	0.20	0.15	0.10																																																																																			
15.0	0.30	0.25	0.20	0.15	0.15	0.15	0.10	0.05																																																																																			
17.5	0.28	0.24	0.20	0.14	0.14	0.14	0.08	0.04																																																																																			
20.0	0.25	0.22	0.18	0.12	0.12	0.12	0.05	0.02																																																																																			
Distribusi bobot akhir	Hidden Layer 1 → Hidden Layer 2																																																																																										

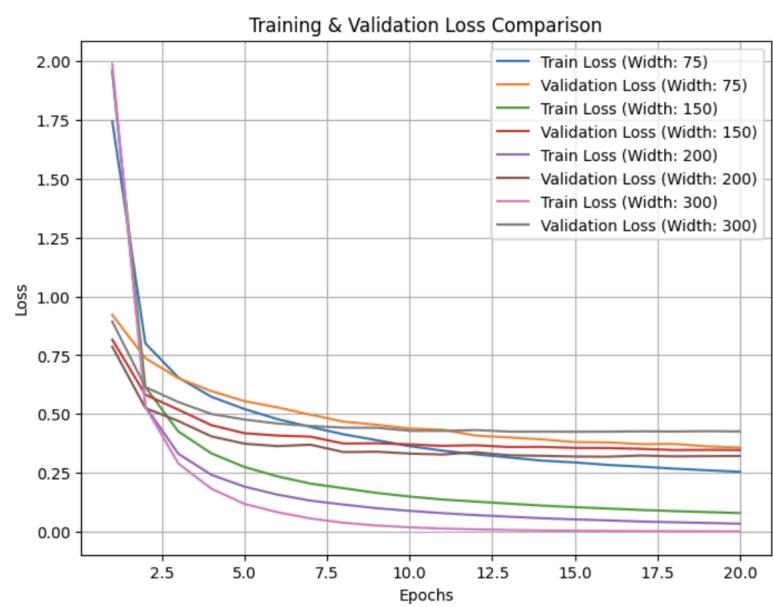


2.2.1.3 Hidden Layer dengan Jumlah Neuron 200

Hasil akhir prediksi		Confusion Matrix												
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9			
		1363	2	6	2	2	6	9	3	5	4	1400	-	
		1	1525	8	2	2	3	0	3	8	4	1200	-	
		2	8	2	1315	8	6	1	1	10	13	4	1000	-
		3	3	0	22	1229	2	33	1	14	22	13	800	-
		4	2	0	15	0	1297	1	13	6	10	51	600	-
		5	5	2	5	28	3	1173	12	7	18	8	400	-
		6	13	4	10	0	12	13	1350	2	13	1	200	-
		7	6	11	18	13	14	5	0	1359	10	27	0	-
		8	8	10	19	26	6	19	4	4	1248	18		
		9	6	1	9	10	40	18	2	27	14	1309		
		0	1	2	3	4	5	6	7	8	9			

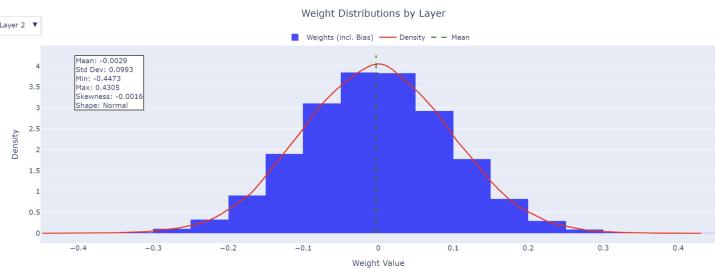
Skor validasi	Weighted F1-Score: 0.9406																																																																						
	Macro F1-Score: 0.9399																																																																						
	Cohen's Kappa: 0.9339																																																																						
	ROC-AUC (Weighted): 0.9968																																																																						
	Validation Log Loss: 0.3222																																																																						
	Classification Report:																																																																						
	<table> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.96</td> <td>0.97</td> <td>0.97</td> <td>1402</td> </tr> <tr> <td>1</td> <td>0.98</td> <td>0.98</td> <td>0.98</td> <td>1556</td> </tr> <tr> <td>2</td> <td>0.92</td> <td>0.96</td> <td>0.94</td> <td>1368</td> </tr> <tr> <td>3</td> <td>0.93</td> <td>0.92</td> <td>0.93</td> <td>1339</td> </tr> <tr> <td>4</td> <td>0.94</td> <td>0.93</td> <td>0.93</td> <td>1395</td> </tr> <tr> <td>5</td> <td>0.92</td> <td>0.93</td> <td>0.93</td> <td>1261</td> </tr> <tr> <td>6</td> <td>0.97</td> <td>0.95</td> <td>0.96</td> <td>1418</td> </tr> <tr> <td>7</td> <td>0.95</td> <td>0.93</td> <td>0.94</td> <td>1463</td> </tr> <tr> <td>8</td> <td>0.92</td> <td>0.92</td> <td>0.92</td> <td>1362</td> </tr> <tr> <td>9</td> <td>0.91</td> <td>0.91</td> <td>0.91</td> <td>1436</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.94</td> <td>14000</td> </tr> <tr> <td>macro avg</td> <td>0.94</td> <td>0.94</td> <td>0.94</td> <td>14000</td> </tr> <tr> <td>weighted avg</td> <td>0.94</td> <td>0.94</td> <td>0.94</td> <td>14000</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.96	0.97	0.97	1402	1	0.98	0.98	0.98	1556	2	0.92	0.96	0.94	1368	3	0.93	0.92	0.93	1339	4	0.94	0.93	0.93	1395	5	0.92	0.93	0.93	1261	6	0.97	0.95	0.96	1418	7	0.95	0.93	0.94	1463	8	0.92	0.92	0.92	1362	9	0.91	0.91	0.91	1436	accuracy			0.94	14000	macro avg	0.94	0.94	0.94	14000	weighted avg	0.94	0.94	0.94	14000
	precision	recall	f1-score	support																																																																			
0	0.96	0.97	0.97	1402																																																																			
1	0.98	0.98	0.98	1556																																																																			
2	0.92	0.96	0.94	1368																																																																			
3	0.93	0.92	0.93	1339																																																																			
4	0.94	0.93	0.93	1395																																																																			
5	0.92	0.93	0.93	1261																																																																			
6	0.97	0.95	0.96	1418																																																																			
7	0.95	0.93	0.94	1463																																																																			
8	0.92	0.92	0.92	1362																																																																			
9	0.91	0.91	0.91	1436																																																																			
accuracy			0.94	14000																																																																			
macro avg	0.94	0.94	0.94	14000																																																																			
weighted avg	0.94	0.94	0.94	14000																																																																			

Grafik loss

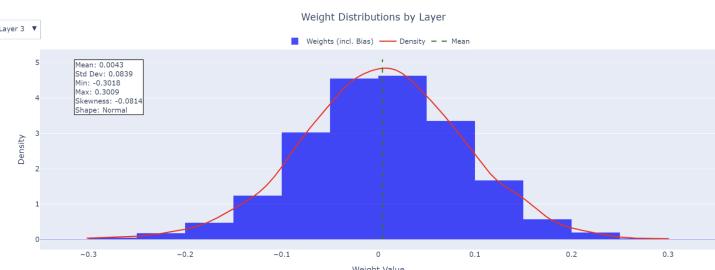


Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

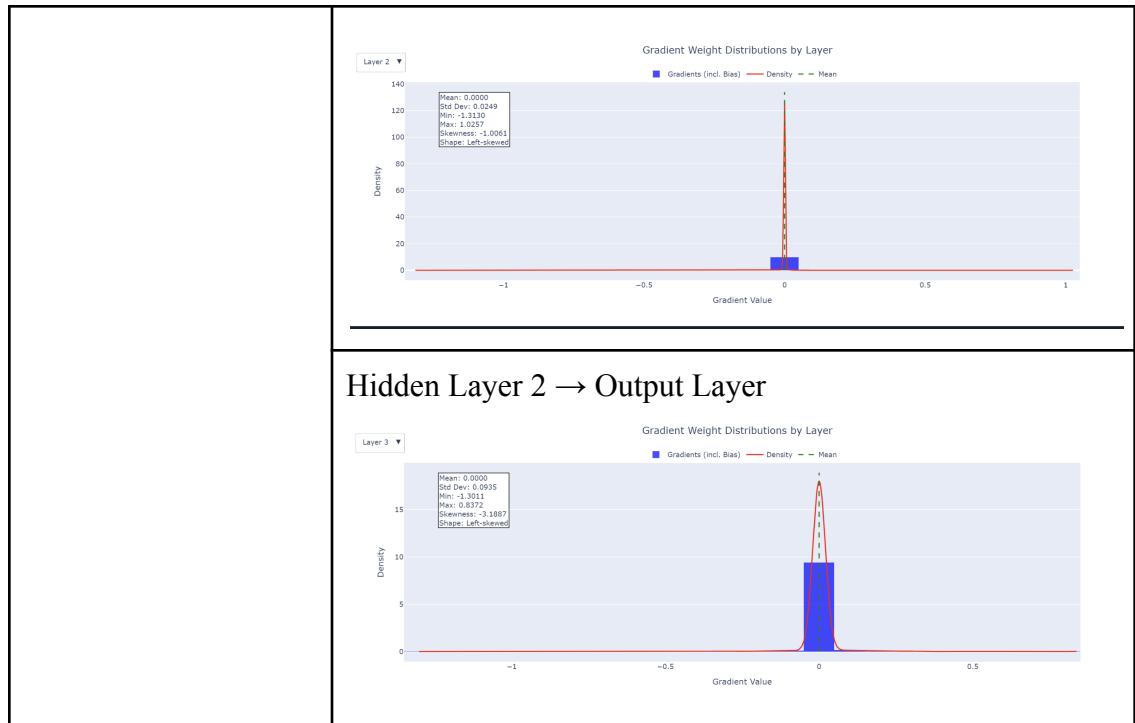


Hidden Layer 2 → Output Layer

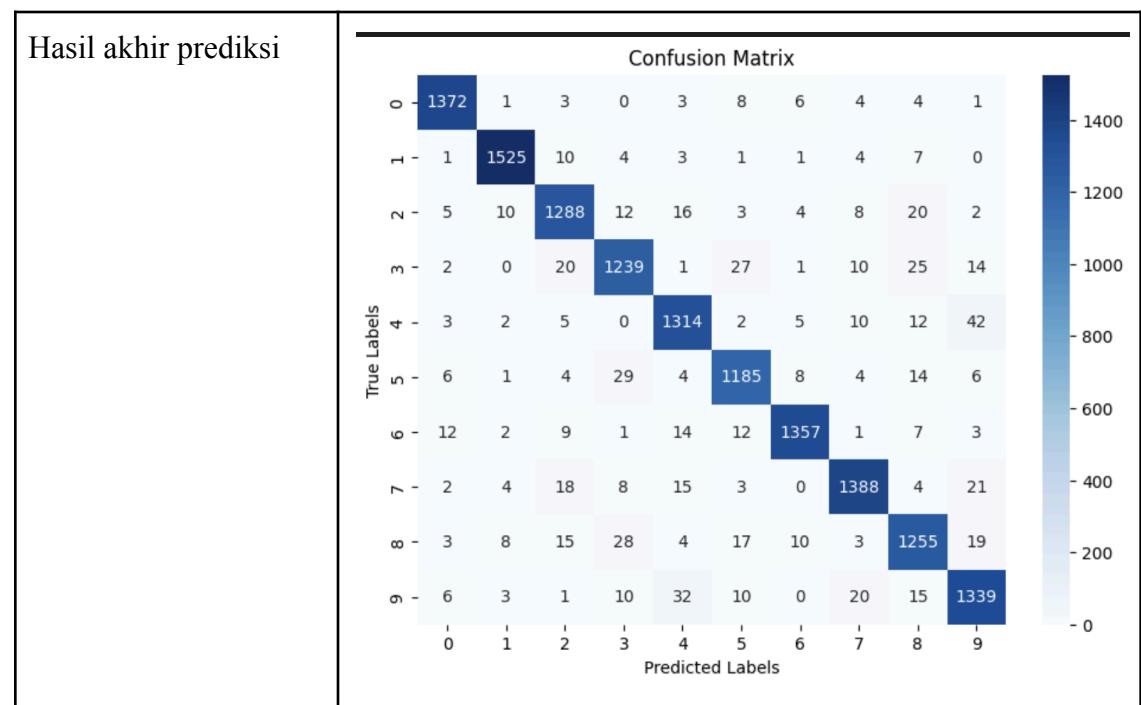


Distribusi gradien bobot akhir

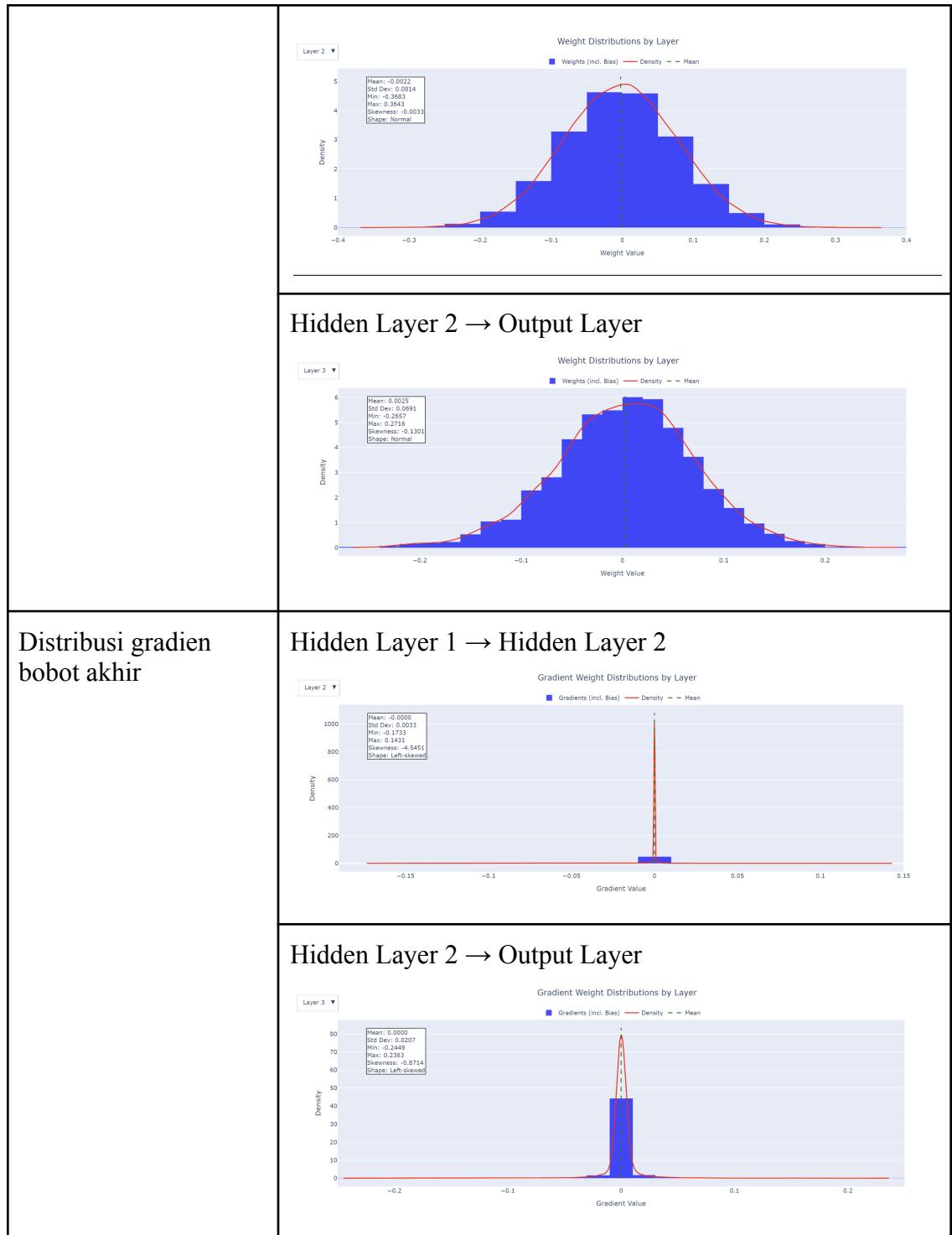
Hidden Layer 1 → Hidden Layer 2



2.2.1.4 Hidden Layer dengan Jumlah Neuron 300



Skor validasi	<p>Weighted F1-Score: 0.9473 Macro F1-Score: 0.9467 Cohen's Kappa: 0.9414 ROC-AUC (Weighted): 0.9972</p> <p>Validation Log Loss: 0.4281</p> <p>Classification Report:</p> <table border="1"> <thead> <tr> <th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr> </thead> <tbody> <tr><td>0</td><td>0.97</td><td>0.98</td><td>0.98</td><td>1402</td></tr> <tr><td>1</td><td>0.98</td><td>0.98</td><td>0.98</td><td>1556</td></tr> <tr><td>2</td><td>0.94</td><td>0.94</td><td>0.94</td><td>1368</td></tr> <tr><td>3</td><td>0.93</td><td>0.93</td><td>0.93</td><td>1339</td></tr> <tr><td>4</td><td>0.93</td><td>0.94</td><td>0.94</td><td>1395</td></tr> <tr><td>5</td><td>0.93</td><td>0.94</td><td>0.94</td><td>1261</td></tr> <tr><td>6</td><td>0.97</td><td>0.96</td><td>0.97</td><td>1418</td></tr> <tr><td>7</td><td>0.96</td><td>0.95</td><td>0.95</td><td>1463</td></tr> <tr><td>8</td><td>0.92</td><td>0.92</td><td>0.92</td><td>1362</td></tr> <tr><td>9</td><td>0.93</td><td>0.93</td><td>0.93</td><td>1436</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.95</td><td>14000</td></tr> <tr><td>macro avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>14000</td></tr> <tr><td>weighted avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>14000</td></tr> </tbody> </table>		precision	recall	f1-score	support	0	0.97	0.98	0.98	1402	1	0.98	0.98	0.98	1556	2	0.94	0.94	0.94	1368	3	0.93	0.93	0.93	1339	4	0.93	0.94	0.94	1395	5	0.93	0.94	0.94	1261	6	0.97	0.96	0.97	1418	7	0.96	0.95	0.95	1463	8	0.92	0.92	0.92	1362	9	0.93	0.93	0.93	1436	accuracy			0.95	14000	macro avg	0.95	0.95	0.95	14000	weighted avg	0.95	0.95	0.95	14000
	precision	recall	f1-score	support																																																																			
0	0.97	0.98	0.98	1402																																																																			
1	0.98	0.98	0.98	1556																																																																			
2	0.94	0.94	0.94	1368																																																																			
3	0.93	0.93	0.93	1339																																																																			
4	0.93	0.94	0.94	1395																																																																			
5	0.93	0.94	0.94	1261																																																																			
6	0.97	0.96	0.97	1418																																																																			
7	0.96	0.95	0.95	1463																																																																			
8	0.92	0.92	0.92	1362																																																																			
9	0.93	0.93	0.93	1436																																																																			
accuracy			0.95	14000																																																																			
macro avg	0.95	0.95	0.95	14000																																																																			
weighted avg	0.95	0.95	0.95	14000																																																																			
Grafik loss	<p>Training & Validation Loss Comparison</p> <p>The graph plots Loss (Y-axis, 0.00 to 2.00) against Epochs (X-axis, 2.5 to 20.0). There are four pairs of lines for each width: Train Loss (solid) and Validation Loss (dashed). The legend indicates four widths: 75, 150, 200, and 300. All lines show a downward trend, indicating decreasing loss over time. The 300-width lines reach the lowest loss most rapidly.</p> <table border="1"> <caption>Approximate Loss Data from Graph</caption> <thead> <tr> <th>Epoch</th> <th>75 Width</th> <th>150 Width</th> <th>200 Width</th> <th>300 Width</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.80</td><td>1.80</td><td>1.80</td><td>2.00</td></tr> <tr><td>2</td><td>0.80</td><td>0.80</td><td>0.80</td><td>0.90</td></tr> <tr><td>5</td><td>0.50</td><td>0.50</td><td>0.50</td><td>0.60</td></tr> <tr><td>10</td><td>0.40</td><td>0.40</td><td>0.40</td><td>0.45</td></tr> <tr><td>20</td><td>0.30</td><td>0.30</td><td>0.30</td><td>0.35</td></tr> </tbody> </table>	Epoch	75 Width	150 Width	200 Width	300 Width	1	1.80	1.80	1.80	2.00	2	0.80	0.80	0.80	0.90	5	0.50	0.50	0.50	0.60	10	0.40	0.40	0.40	0.45	20	0.30	0.30	0.30	0.35																																								
Epoch	75 Width	150 Width	200 Width	300 Width																																																																			
1	1.80	1.80	1.80	2.00																																																																			
2	0.80	0.80	0.80	0.90																																																																			
5	0.50	0.50	0.50	0.60																																																																			
10	0.40	0.40	0.40	0.45																																																																			
20	0.30	0.30	0.30	0.35																																																																			
Distribusi bobot akhir	Hidden Layer 1 → Hidden Layer 2																																																																						



Pada pengujian model FFNN dengan variasi *width* (lebar), didapatkan bahwa jumlah neuron sebanyak 300 neuron memberikan performa terbaik untuk model. Hal ini ditandai dengan semakin membesarnya loss jika dilakukan perubahan jumlah neuron dari 300 neuron. Semakin kecil jumlah neuron, besar loss yang model miliki pada proses training

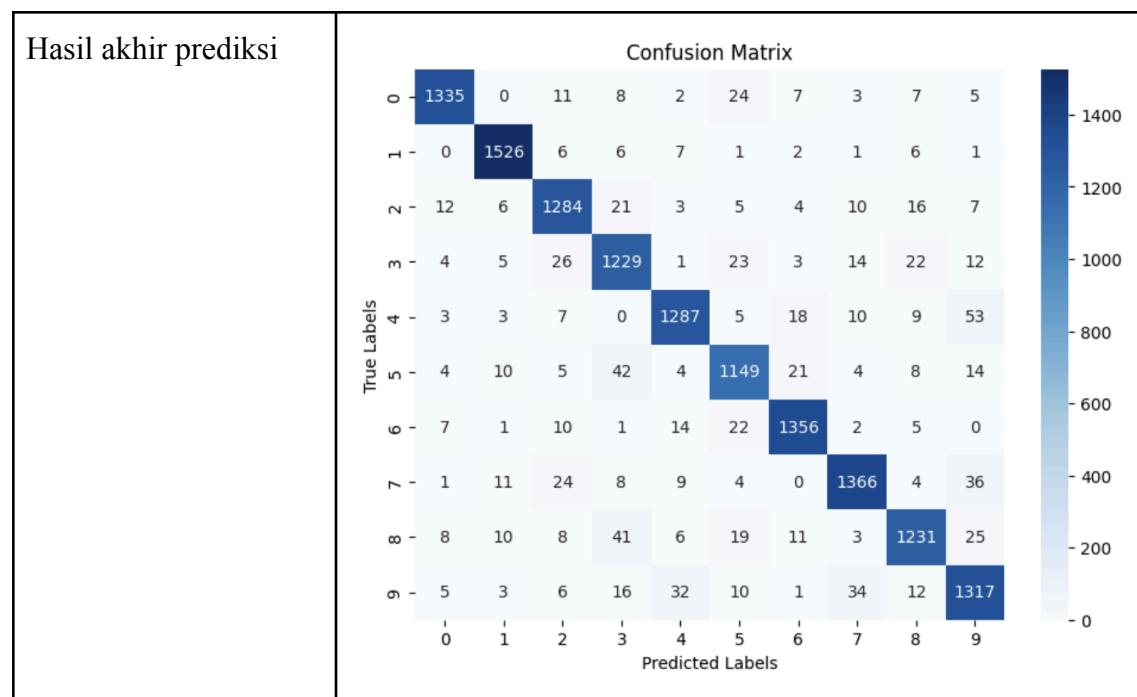
semakin membesar sehingga menyebabkan akurasi hasil pelatihan yang lebih kecil. Dari hasil pengamatan ini, dapat disimpulkan bahwa

2.2.2. Pengaruh Depth

Parameter kontrol yang digunakan untuk membandingkan pengaruh antara metode inisialisasi bobot yang diimplementasikan adalah :

Inisialisasi Bobot	:	Normal He Initialization
Fungsi Aktivasi	:	ReLU
Fungsi loss	:	Categorical Cross-Entropy
Batch size	:	64
Learning rate	:	0.001
Epoch maksimal	:	50
Fungsi Aktivasi untuk layer terakhir adalah softmax		

2.2.2.1 Model dengan 2 Hidden Layer



Skor validasi

Weighted F1-Score: 0.9343

Macro F1-Score: 0.9335

Cohen's Kappa: 0.9270

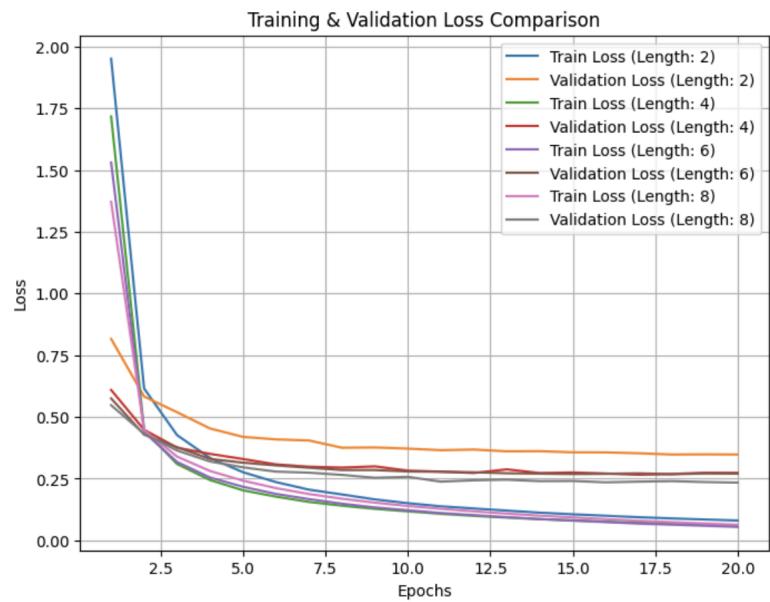
ROC-AUC (Weighted): 0.9951

Validation Log Loss: 0.3482

Classification Report:

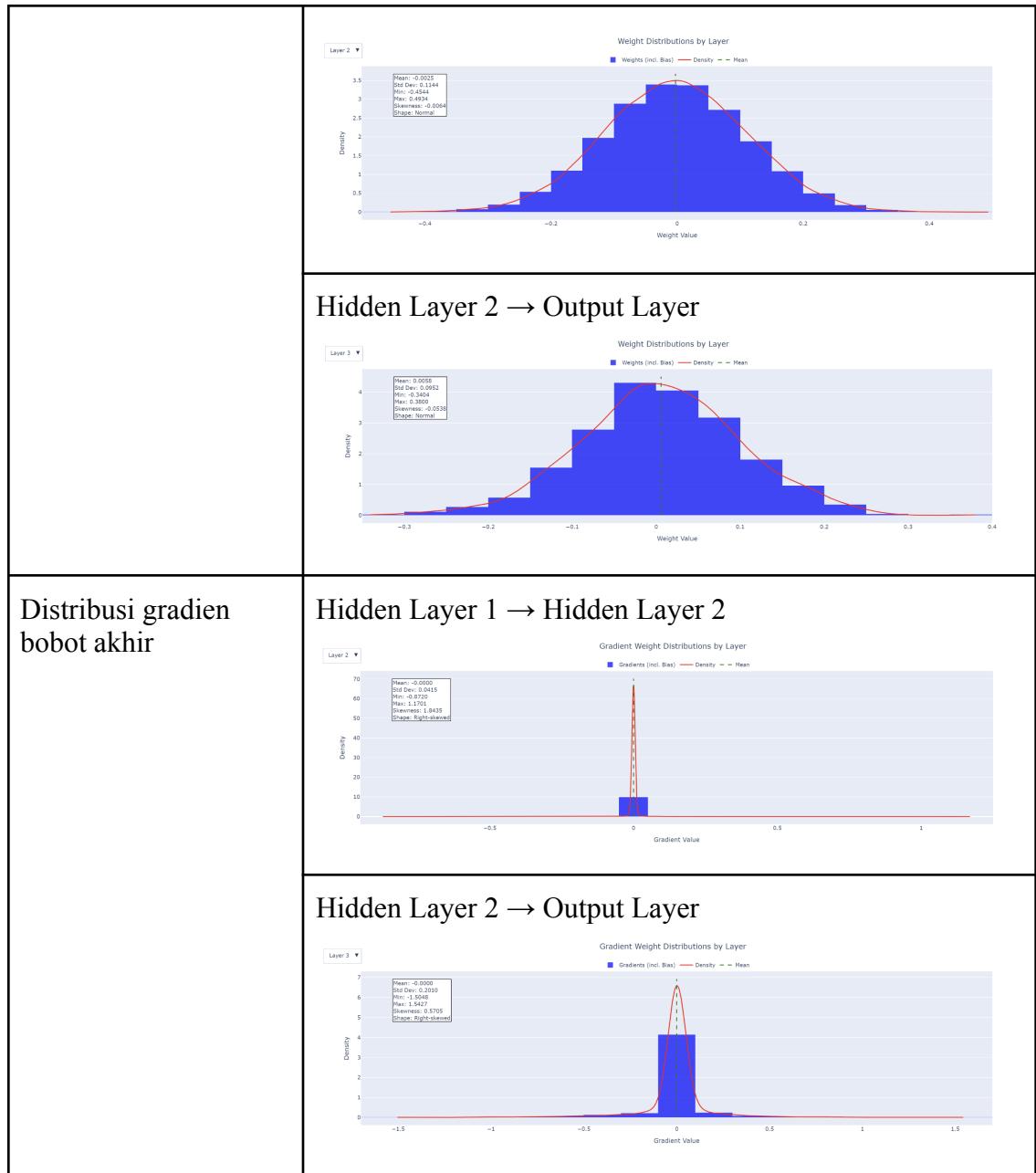
	precision	recall	f1-score	support
0	0.97	0.95	0.96	1402
1	0.97	0.98	0.97	1556
2	0.93	0.94	0.93	1368
3	0.90	0.92	0.91	1339
4	0.94	0.92	0.93	1395
5	0.91	0.91	0.91	1261
6	0.95	0.96	0.95	1418
7	0.94	0.93	0.94	1463
8	0.93	0.90	0.92	1362
9	0.90	0.92	0.91	1436
accuracy			0.93	14000
macro avg	0.93	0.93	0.93	14000
weighted avg	0.93	0.93	0.93	14000

Grafik loss



Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

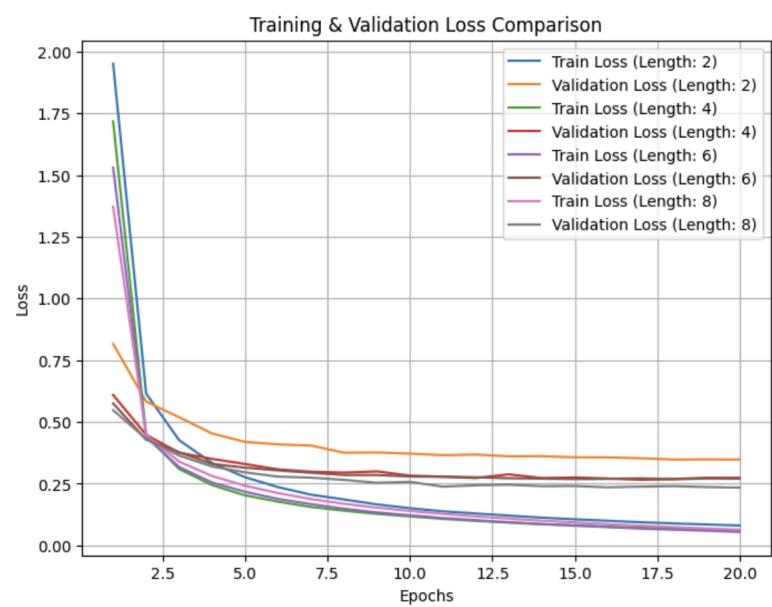


2.2.2.2 Model dengan 4 Hidden Layer

Hasil akhir prediksi		Confusion Matrix											
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9		
		1363	1	10	1	2	12	4	0	7	2	- 1400	
		0	1535	7	4	1	0	1	4	3	1	- 1200	
		8	5	1306	18	11	2	3	6	8	1	- 1000	
		4	5	29	1224	2	29	1	10	27	8	- 800	
		7	0	13	1	1305	5	8	9	11	36	- 600	
		6	5	6	35	6	1162	14	1	15	11	- 400	
		18	3	14	4	15	23	1335	0	3	3	- 200	
		6	7	18	12	16	9	0	1357	7	31	- 0	
		4	15	16	26	5	19	4	0	1260	13		
		9	4	1	26	40	19	0	20	20	1297		
		0	1	2	3	4	5	6	7	8	9		

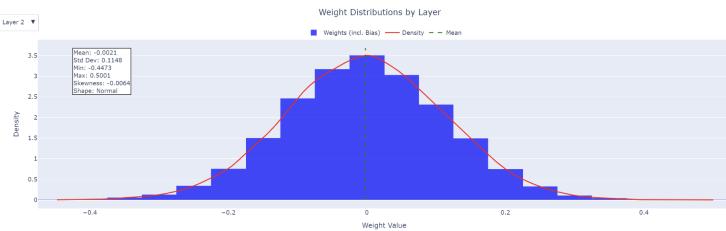
Skor validasi	Weighted F1-Score: 0.9389 Macro F1-Score: 0.9380 Cohen's Kappa: 0.9320 ROC-AUC (Weighted): 0.9965																																																																						
	Validation Log Loss: 0.2734																																																																						
	Classification Report: <table> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.96</td> <td>0.97</td> <td>0.96</td> <td>1402</td> </tr> <tr> <td>1</td> <td>0.97</td> <td>0.99</td> <td>0.98</td> <td>1556</td> </tr> <tr> <td>2</td> <td>0.92</td> <td>0.95</td> <td>0.94</td> <td>1368</td> </tr> <tr> <td>3</td> <td>0.91</td> <td>0.91</td> <td>0.91</td> <td>1339</td> </tr> <tr> <td>4</td> <td>0.93</td> <td>0.94</td> <td>0.93</td> <td>1395</td> </tr> <tr> <td>5</td> <td>0.91</td> <td>0.92</td> <td>0.91</td> <td>1261</td> </tr> <tr> <td>6</td> <td>0.97</td> <td>0.94</td> <td>0.96</td> <td>1418</td> </tr> <tr> <td>7</td> <td>0.96</td> <td>0.93</td> <td>0.95</td> <td>1463</td> </tr> <tr> <td>8</td> <td>0.93</td> <td>0.93</td> <td>0.93</td> <td>1362</td> </tr> <tr> <td>9</td> <td>0.92</td> <td>0.90</td> <td>0.91</td> <td>1436</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.94</td> <td>14000</td> </tr> <tr> <td>macro avg</td> <td>0.94</td> <td>0.94</td> <td>0.94</td> <td>14000</td> </tr> <tr> <td>weighted avg</td> <td>0.94</td> <td>0.94</td> <td>0.94</td> <td>14000</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.96	0.97	0.96	1402	1	0.97	0.99	0.98	1556	2	0.92	0.95	0.94	1368	3	0.91	0.91	0.91	1339	4	0.93	0.94	0.93	1395	5	0.91	0.92	0.91	1261	6	0.97	0.94	0.96	1418	7	0.96	0.93	0.95	1463	8	0.93	0.93	0.93	1362	9	0.92	0.90	0.91	1436	accuracy			0.94	14000	macro avg	0.94	0.94	0.94	14000	weighted avg	0.94	0.94	0.94	14000
	precision	recall	f1-score	support																																																																			
0	0.96	0.97	0.96	1402																																																																			
1	0.97	0.99	0.98	1556																																																																			
2	0.92	0.95	0.94	1368																																																																			
3	0.91	0.91	0.91	1339																																																																			
4	0.93	0.94	0.93	1395																																																																			
5	0.91	0.92	0.91	1261																																																																			
6	0.97	0.94	0.96	1418																																																																			
7	0.96	0.93	0.95	1463																																																																			
8	0.93	0.93	0.93	1362																																																																			
9	0.92	0.90	0.91	1436																																																																			
accuracy			0.94	14000																																																																			
macro avg	0.94	0.94	0.94	14000																																																																			
weighted avg	0.94	0.94	0.94	14000																																																																			

Grafik loss

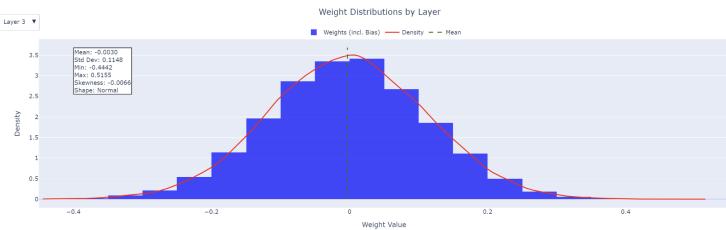


Distribusi bobot akhir

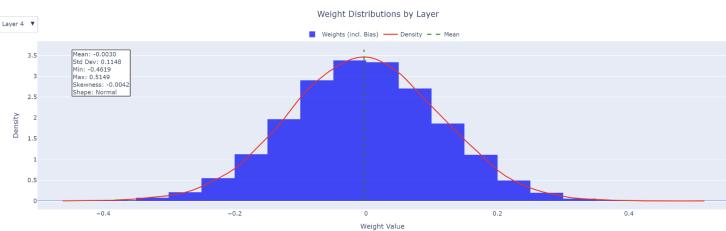
Hidden Layer 1 → Hidden Layer 2



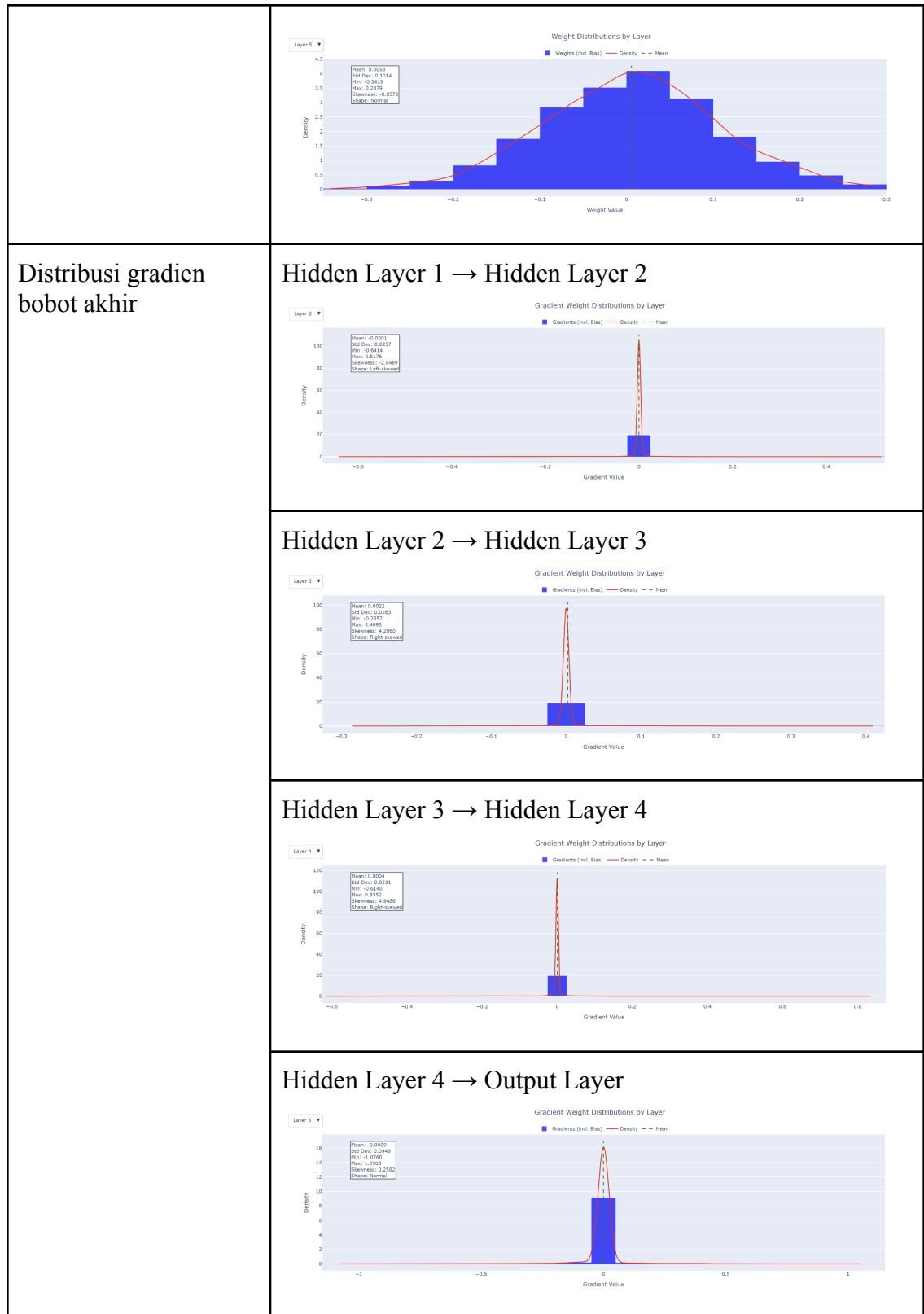
Hidden Layer 2 → Hidden Layer 3



Hidden Layer 3 → Hidden Layer 4



Hidden Layer 4 → Output Layer

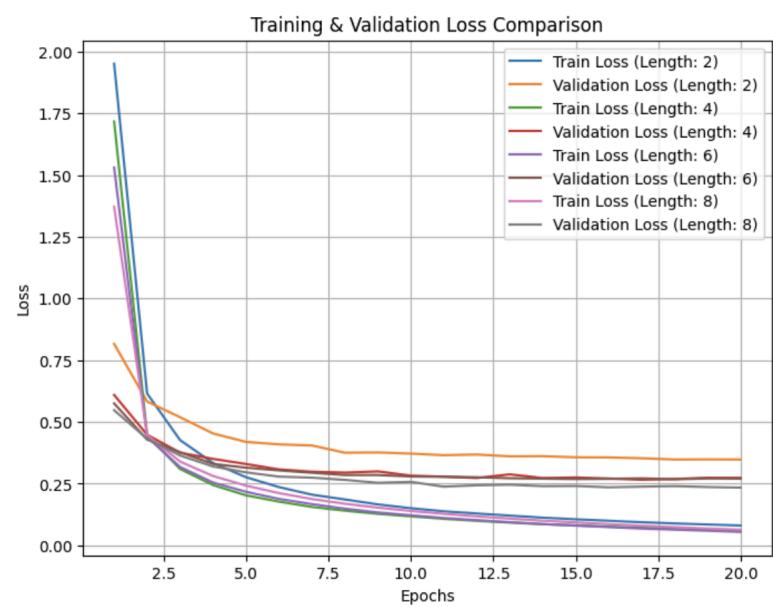


2.2.2.3 Model dengan 6 Hidden Layer

Hasil akhir prediksi		Confusion Matrix										
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9	0
		1342	0	6	5	2	22	11	3	9	2	- 1400
		1	1515	11	8	4	1	1	7	8	1	- 1200
		2	5	7	1293	5	5	5	10	16	21	- 1000
		3	0	4	22	1204	0	36	4	15	46	- 800
		4	0	2	10	0	1299	4	14	9	9	- 600
		5	4	2	5	26	4	1169	13	2	28	- 400
		6	9	2	15	1	11	21	1347	0	10	- 200
		7	2	11	18	6	12	9	1	1354	7	- 0
		8	1	6	7	27	5	13	6	3	1281	- 0
		9	6	3	3	19	51	10	0	30	15	- 0
		0	1	2	3	4	5	6	7	8	9	- 0

Skor validasi	Weighted F1-Score: 0.9360			
	Macro F1-Score: 0.9353			
	Cohen's Kappa: 0.9288			
	ROC-AUC (Weighted): 0.9963			
	Validation Log Loss: 0.2702			
	Classification Report:			
	precision recall f1-score support			
0	0.98	0.96	0.97	1402
1	0.98	0.97	0.97	1556
2	0.93	0.95	0.94	1368
3	0.93	0.90	0.91	1339
4	0.93	0.93	0.93	1395
5	0.91	0.93	0.92	1261
6	0.96	0.95	0.95	1418
7	0.94	0.93	0.93	1463
8	0.89	0.94	0.92	1362
9	0.91	0.90	0.91	1436
	accuracy			0.94
	macro avg	0.94	0.94	0.94
	weighted avg	0.94	0.94	0.94

Grafik loss



Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2



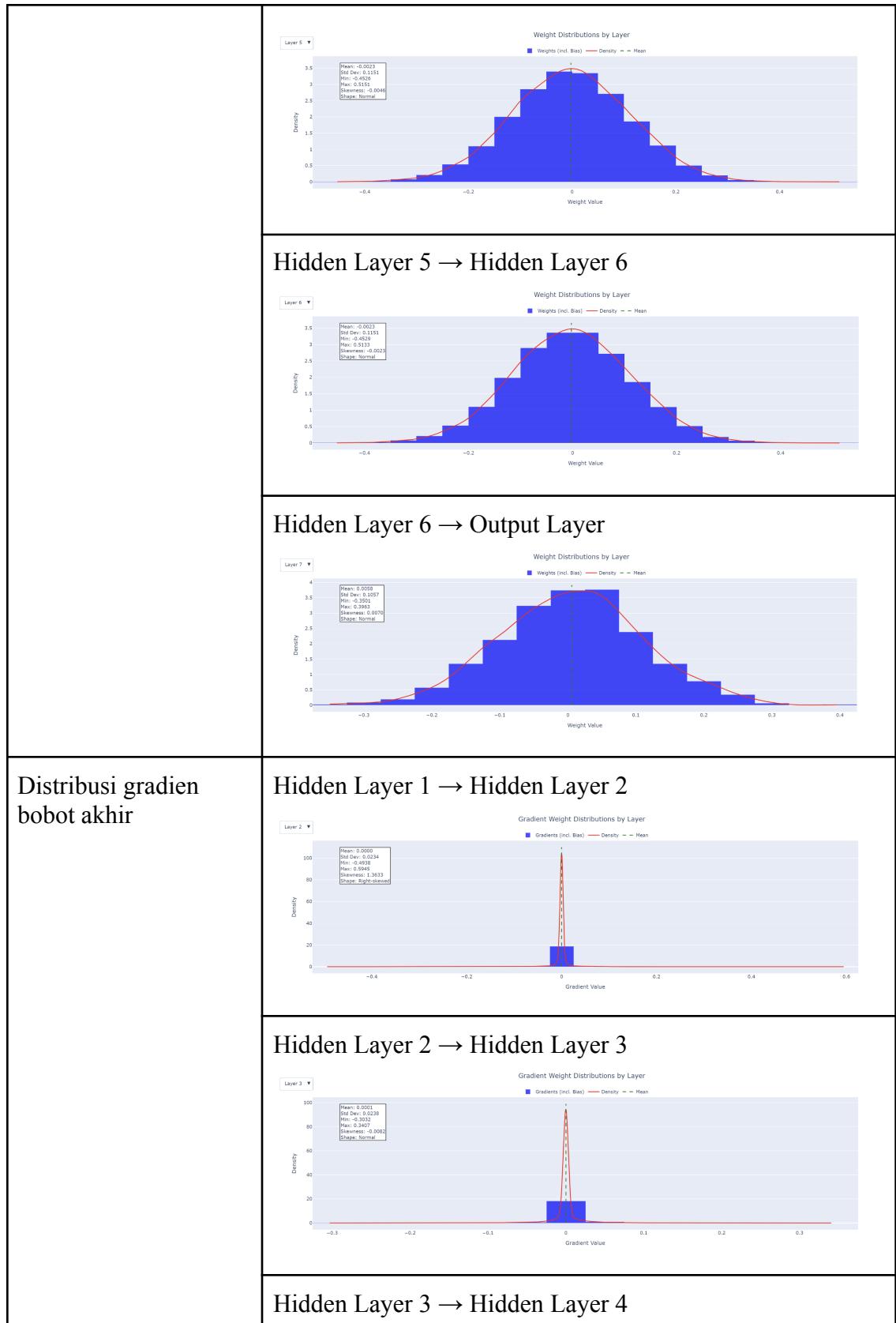
Hidden Layer 2 → Hidden Layer 3

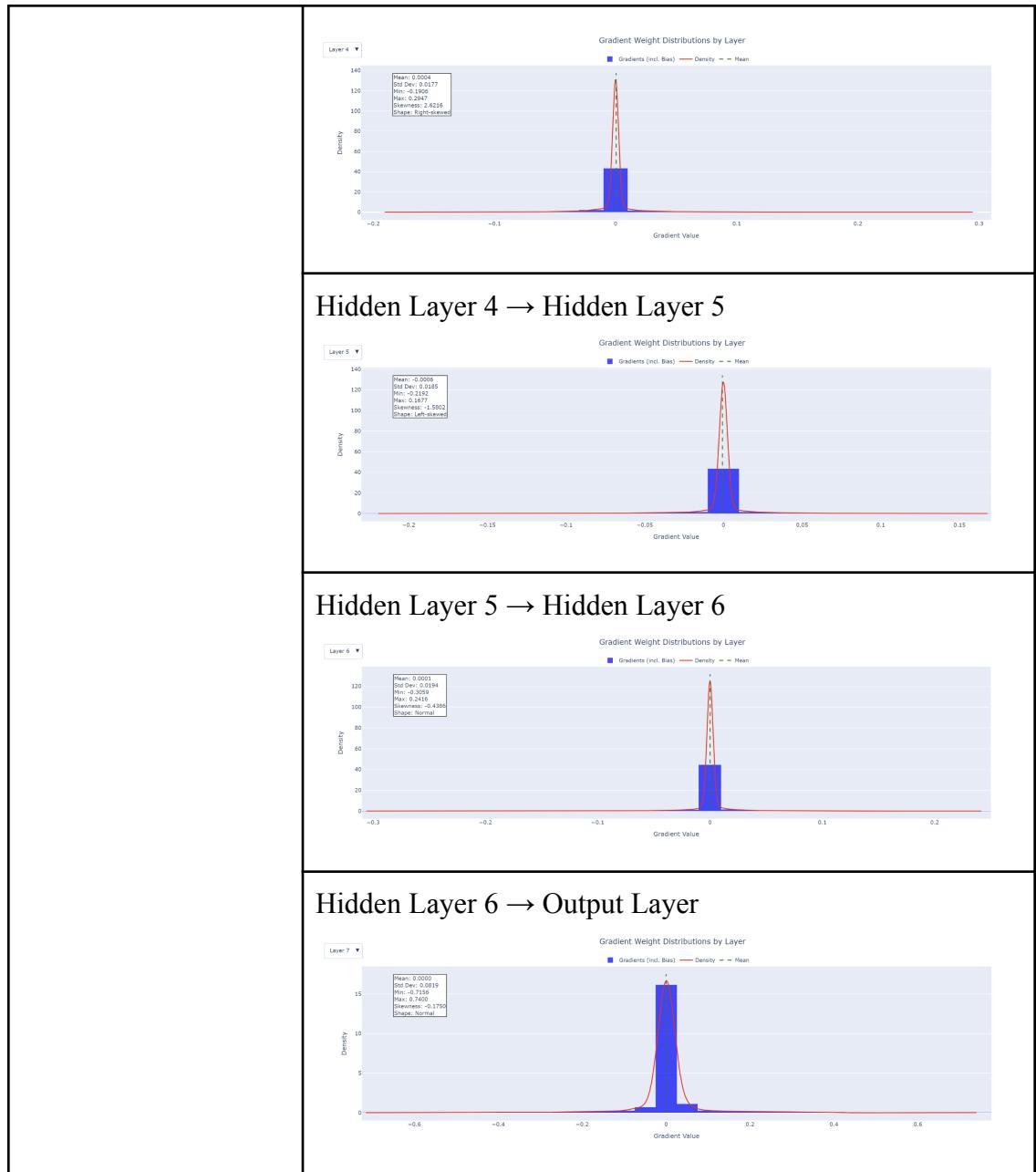


Hidden Layer 3 → Hidden Layer 4



Hidden Layer 4 → Hidden Layer 5

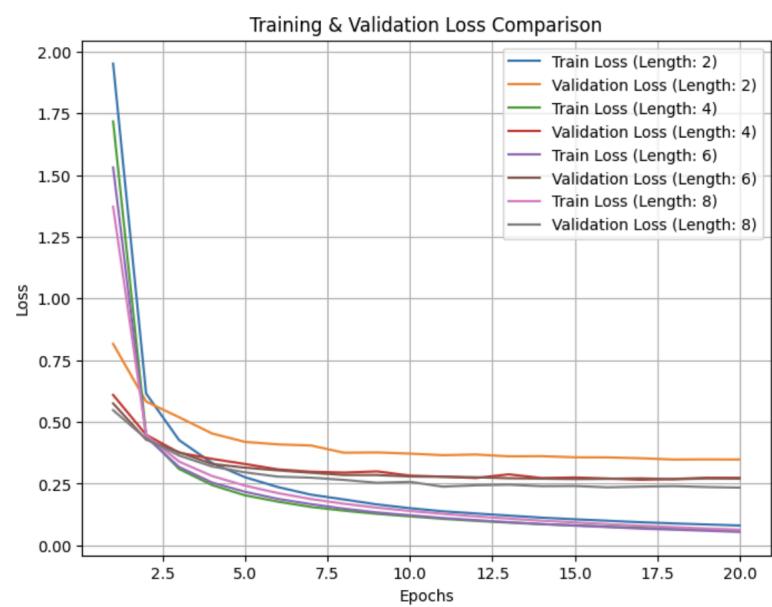




2.2.2.4 Model dengan 8 Hidden Layer

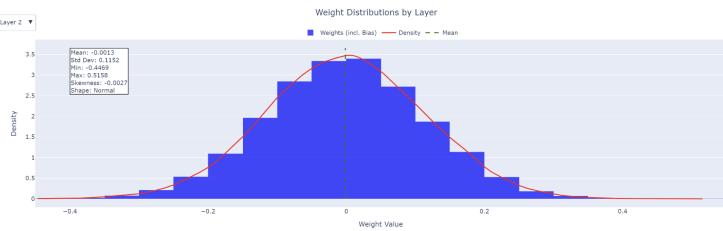
Hasil akhir prediksi		Confusion Matrix																																																																																
		0	1349	1	4	2	0	15	6	9	9	7																																																																						
		1	0	1523	16	2	2	3	0	2	5	3																																																																						
		2	7	2	1300	11	10	5	5	10	13	5																																																																						
		3	1	5	23	1226	1	30	3	7	35	8																																																																						
True Labels	4	-	1	1	12	1	1308	5	9	6	5	47																																																																						
	5	-	6	2	4	29	2	1182	16	4	9	7																																																																						
	6	-	13	1	11	1	8	24	1353	2	5	0																																																																						
	7	-	4	6	15	4	8	3	1	1389	7	26																																																																						
	8	-	3	8	10	35	3	29	6	1	1259	8																																																																						
	9	-	6	4	2	21	37	18	0	36	10	1302																																																																						
		0	1	2	3	4	5	6	7	8	9	9																																																																						
		Predicted Labels																																																																																
Skor validasi		Weighted F1-Score: 0.9423 Macro F1-Score: 0.9415 Cohen's Kappa: 0.9358 ROC-AUC (Weighted): 0.9968 Validation Log Loss: 0.2339 Classification Report: <table> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr><td>0</td><td>0.97</td><td>0.96</td><td>0.97</td><td>1402</td></tr> <tr><td>1</td><td>0.98</td><td>0.98</td><td>0.98</td><td>1556</td></tr> <tr><td>2</td><td>0.93</td><td>0.95</td><td>0.94</td><td>1368</td></tr> <tr><td>3</td><td>0.92</td><td>0.92</td><td>0.92</td><td>1339</td></tr> <tr><td>4</td><td>0.95</td><td>0.94</td><td>0.94</td><td>1395</td></tr> <tr><td>5</td><td>0.90</td><td>0.94</td><td>0.92</td><td>1261</td></tr> <tr><td>6</td><td>0.97</td><td>0.95</td><td>0.96</td><td>1418</td></tr> <tr><td>7</td><td>0.95</td><td>0.95</td><td>0.95</td><td>1463</td></tr> <tr><td>8</td><td>0.93</td><td>0.92</td><td>0.93</td><td>1362</td></tr> <tr><td>9</td><td>0.92</td><td>0.91</td><td>0.91</td><td>1436</td></tr> <tr><td>accuracy</td><td></td><td></td><td>0.94</td><td>14000</td></tr> <tr><td>macro avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>14000</td></tr> <tr><td>weighted avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>14000</td></tr> </tbody> </table>												precision	recall	f1-score	support	0	0.97	0.96	0.97	1402	1	0.98	0.98	0.98	1556	2	0.93	0.95	0.94	1368	3	0.92	0.92	0.92	1339	4	0.95	0.94	0.94	1395	5	0.90	0.94	0.92	1261	6	0.97	0.95	0.96	1418	7	0.95	0.95	0.95	1463	8	0.93	0.92	0.93	1362	9	0.92	0.91	0.91	1436	accuracy			0.94	14000	macro avg	0.94	0.94	0.94	14000	weighted avg	0.94	0.94	0.94	14000
	precision	recall	f1-score	support																																																																														
0	0.97	0.96	0.97	1402																																																																														
1	0.98	0.98	0.98	1556																																																																														
2	0.93	0.95	0.94	1368																																																																														
3	0.92	0.92	0.92	1339																																																																														
4	0.95	0.94	0.94	1395																																																																														
5	0.90	0.94	0.92	1261																																																																														
6	0.97	0.95	0.96	1418																																																																														
7	0.95	0.95	0.95	1463																																																																														
8	0.93	0.92	0.93	1362																																																																														
9	0.92	0.91	0.91	1436																																																																														
accuracy			0.94	14000																																																																														
macro avg	0.94	0.94	0.94	14000																																																																														
weighted avg	0.94	0.94	0.94	14000																																																																														

Grafik loss



Distribusi bobot akhir

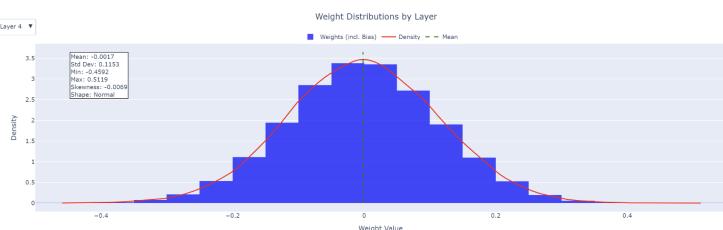
Hidden Layer 1 → Hidden Layer 2



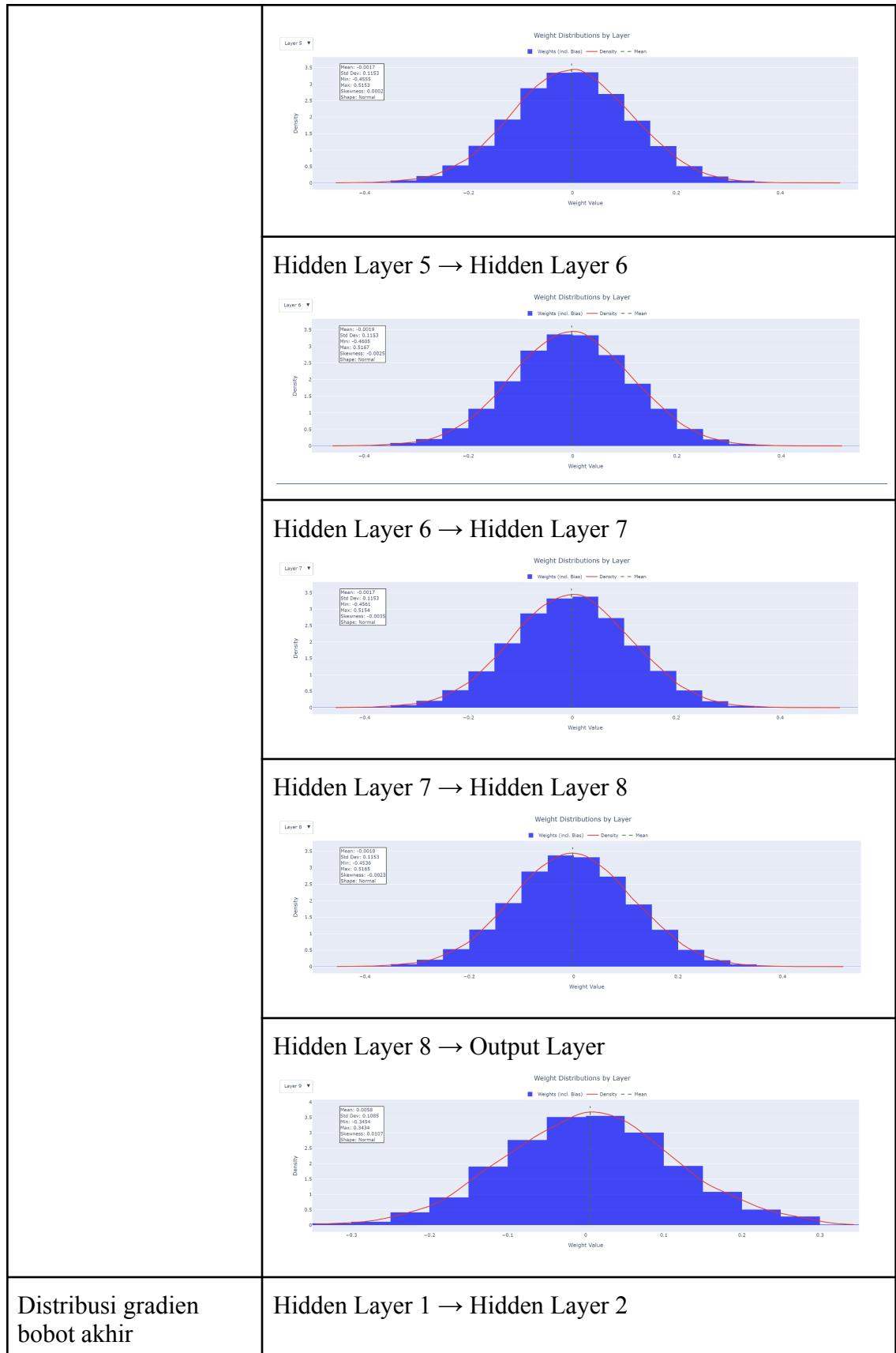
Hidden Layer 2 → Hidden Layer 3

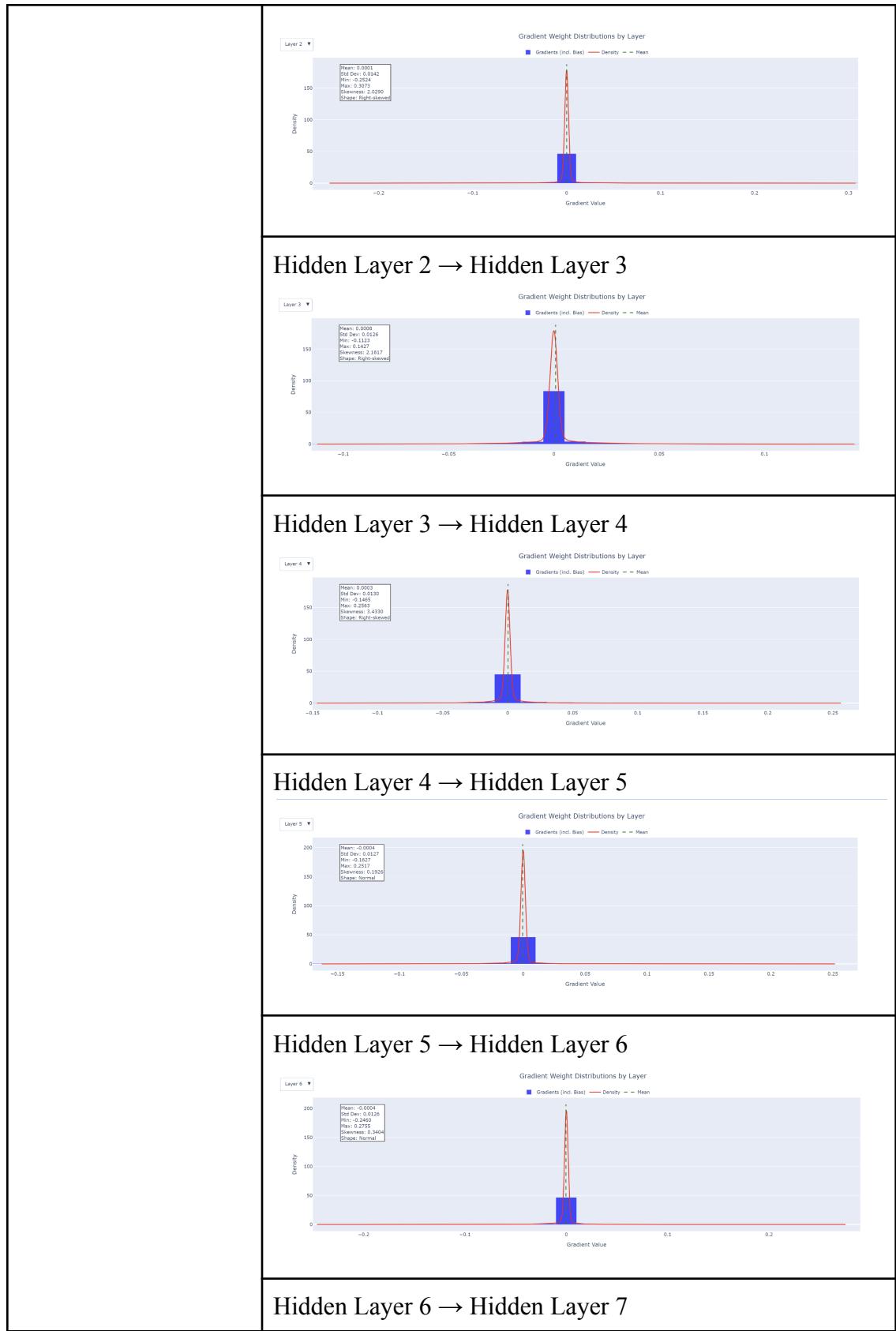


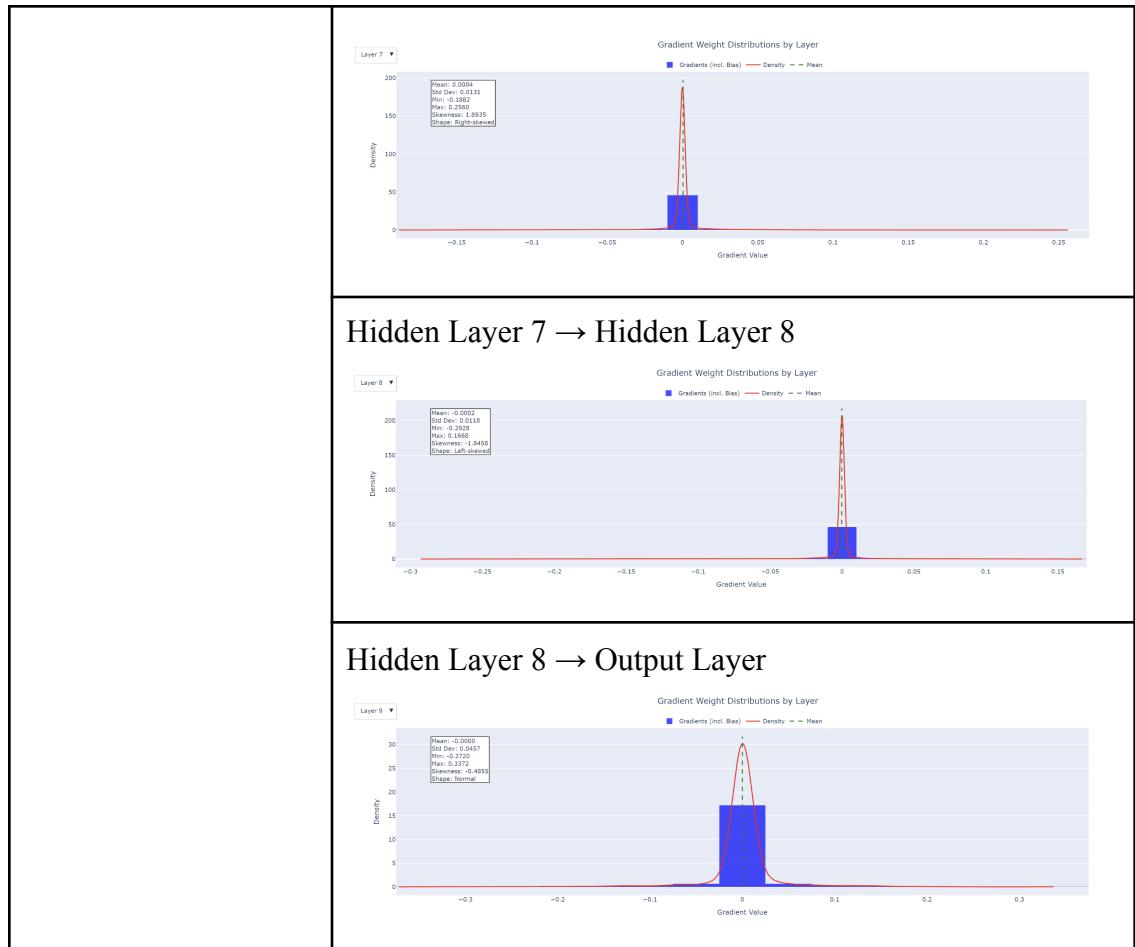
Hidden Layer 3 → Hidden Layer 4



Hidden Layer 4 → Hidden Layer 5







Pada pengujian model FFNN dengan variasi *depth* (kedalaman), didapatkan bahwa jumlah hidden layer tertinggi memberikan performa terbaik untuk model. Hal ini ditandai dengan semakin membesarnya loss hasil pelatihan jika dibandingkan dengan jumlah hidden layer yang lebih kecil.

2.2.3. Pengaruh Fungsi Aktivasi

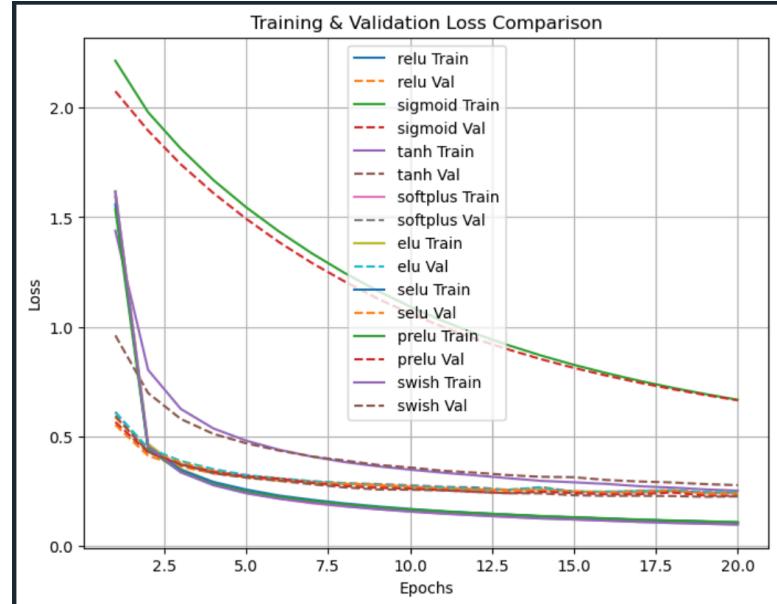
Parameter kontrol yang digunakan untuk membandingkan pengaruh antara fungsi aktivasi yang diimplementasikan adalah :

Jumlah layer	: 4 (input layer, hidden layer 1, hidden layer 2, output layer)
Neuron tiap layer	: 784 neuron, 150 neuron, 75 neuron, 10 neuron
Inisialisasi Bobot	: Normal He Initialization
Fungsi loss	: Categorical Cross-Entropy
Batch size	: 64
Learning rate	: 0.001
Epoch maksimal	: 20
Fungsi Aktivasi untuk layer terakhir	: softmax

2.2.2.1 Fungsi Aktivasi sigmoid

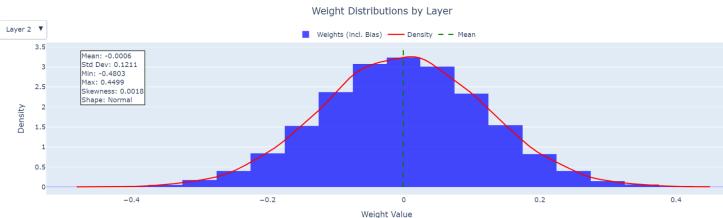
<p>Hasil akhir prediksi</p>	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="12">Confusion Matrix</th> </tr> <tr> <th rowspan="2">True Labels</th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> <th>0</th> </tr> </thead> <tbody> <tr> <th>0</th> <td>1358</td> <td>0</td> <td>6</td> <td>1</td> <td>0</td> <td>10</td> <td>14</td> <td>1</td> <td>12</td> <td>0</td> <td>1400</td> </tr> <tr> <th>1</th> <td>0</td> <td>1515</td> <td>12</td> <td>4</td> <td>1</td> <td>2</td> <td>6</td> <td>3</td> <td>12</td> <td>1</td> <td>1200</td> </tr> <tr> <th>2</th> <td>17</td> <td>11</td> <td>1199</td> <td>18</td> <td>31</td> <td>1</td> <td>35</td> <td>26</td> <td>27</td> <td>3</td> <td>1000</td> </tr> <tr> <th>3</th> <td>15</td> <td>6</td> <td>50</td> <td>1125</td> <td>2</td> <td>35</td> <td>18</td> <td>30</td> <td>41</td> <td>17</td> <td>800</td> </tr> <tr> <th>4</th> <td>1</td> <td>4</td> <td>9</td> <td>0</td> <td>1256</td> <td>1</td> <td>26</td> <td>4</td> <td>11</td> <td>83</td> <td>600</td> </tr> <tr> <th>5</th> <td>63</td> <td>12</td> <td>10</td> <td>154</td> <td>20</td> <td>900</td> <td>25</td> <td>9</td> <td>54</td> <td>14</td> <td>400</td> </tr> <tr> <th>6</th> <td>27</td> <td>8</td> <td>26</td> <td>1</td> <td>15</td> <td>26</td> <td>1310</td> <td>1</td> <td>4</td> <td>0</td> <td>200</td> </tr> <tr> <th>7</th> <td>17</td> <td>32</td> <td>23</td> <td>2</td> <td>29</td> <td>1</td> <td>3</td> <td>1301</td> <td>4</td> <td>51</td> <td>0</td> </tr> <tr> <th>8</th> <td>9</td> <td>28</td> <td>35</td> <td>101</td> <td>6</td> <td>51</td> <td>21</td> <td>12</td> <td>1058</td> <td>41</td> <td>-</td> </tr> <tr> <th>9</th> <td>12</td> <td>4</td> <td>16</td> <td>28</td> <td>73</td> <td>12</td> <td>4</td> <td>91</td> <td>15</td> <td>1181</td> <td>-</td> </tr> <tr> <th>0</th> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>9</td> <td>-</td> </tr> </tbody> </table>	Confusion Matrix												True Labels	0	1	2	3	4	5	6	7	8	9	0	0	1358	0	6	1	0	10	14	1	12	0	1400	1	0	1515	12	4	1	2	6	3	12	1	1200	2	17	11	1199	18	31	1	35	26	27	3	1000	3	15	6	50	1125	2	35	18	30	41	17	800	4	1	4	9	0	1256	1	26	4	11	83	600	5	63	12	10	154	20	900	25	9	54	14	400	6	27	8	26	1	15	26	1310	1	4	0	200	7	17	32	23	2	29	1	3	1301	4	51	0	8	9	28	35	101	6	51	21	12	1058	41	-	9	12	4	16	28	73	12	4	91	15	1181	-	0	1	2	3	4	5	6	7	8	9	9	-
Confusion Matrix																																																																																																																																																													
True Labels	0	1	2	3	4	5	6	7	8	9	0																																																																																																																																																		
	0	1358	0	6	1	0	10	14	1	12	0	1400																																																																																																																																																	
1	0	1515	12	4	1	2	6	3	12	1	1200																																																																																																																																																		
2	17	11	1199	18	31	1	35	26	27	3	1000																																																																																																																																																		
3	15	6	50	1125	2	35	18	30	41	17	800																																																																																																																																																		
4	1	4	9	0	1256	1	26	4	11	83	600																																																																																																																																																		
5	63	12	10	154	20	900	25	9	54	14	400																																																																																																																																																		
6	27	8	26	1	15	26	1310	1	4	0	200																																																																																																																																																		
7	17	32	23	2	29	1	3	1301	4	51	0																																																																																																																																																		
8	9	28	35	101	6	51	21	12	1058	41	-																																																																																																																																																		
9	12	4	16	28	73	12	4	91	15	1181	-																																																																																																																																																		
0	1	2	3	4	5	6	7	8	9	9	-																																																																																																																																																		
<p>Skor validasi</p>	<pre> Weighted F1-Score: 0.8703 Macro F1-Score: 0.8681 Cohen's Kappa: 0.8573 ROC-AUC (Weighted): 0.9852 Validation Log Loss: 0.6654 Classification Report: precision recall f1-score support 0 0.89 0.97 0.93 1402 1 0.94 0.97 0.95 1556 2 0.87 0.88 0.87 1368 3 0.78 0.84 0.81 1339 4 0.88 0.90 0.89 1395 5 0.87 0.71 0.78 1261 6 0.90 0.92 0.91 1418 7 0.88 0.89 0.88 1463 8 0.85 0.78 0.81 1362 9 0.85 0.82 0.84 1436 accuracy 0.87 14000 macro avg 0.87 0.87 0.87 14000 weighted avg 0.87 0.87 0.87 14000 </pre>																																																																																																																																																												

Grafik loss

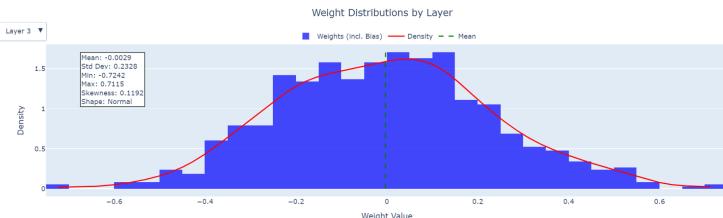


Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

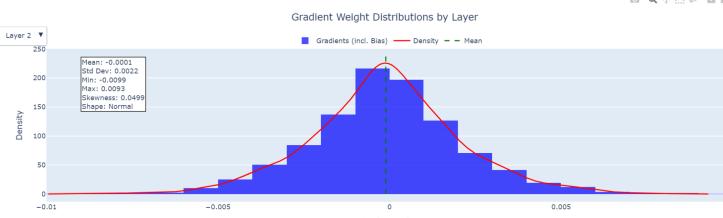


Hidden Layer 2 → Output Layer

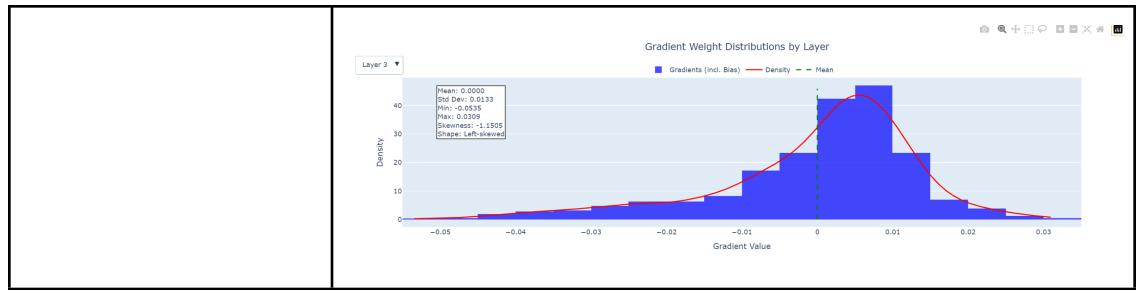


Distribusi gradien bobot akhir

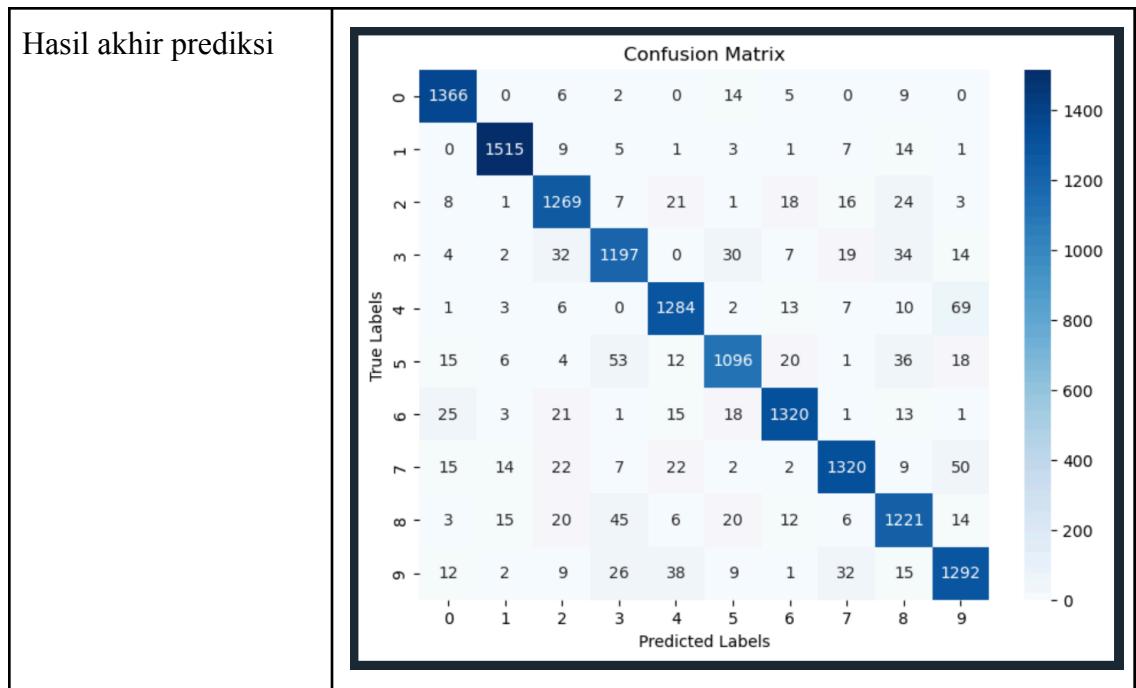
Hidden Layer 1 → Hidden Layer 2



Hidden Layer 2 → Output Layer



2.2.2.2 Fungsi Aktivasi tanh



Skor validasi

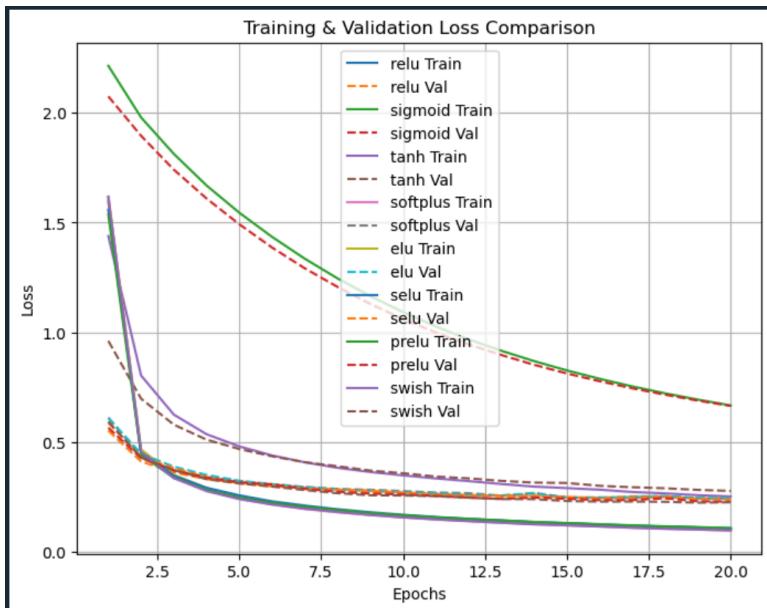
```
Weighted F1-Score: 0.9199  
Macro F1-Score: 0.9189  
Cohen's Kappa: 0.9111  
ROC-AUC (Weighted): 0.9946
```

```
Validation Log Loss: 0.2787
```

Classification Report:

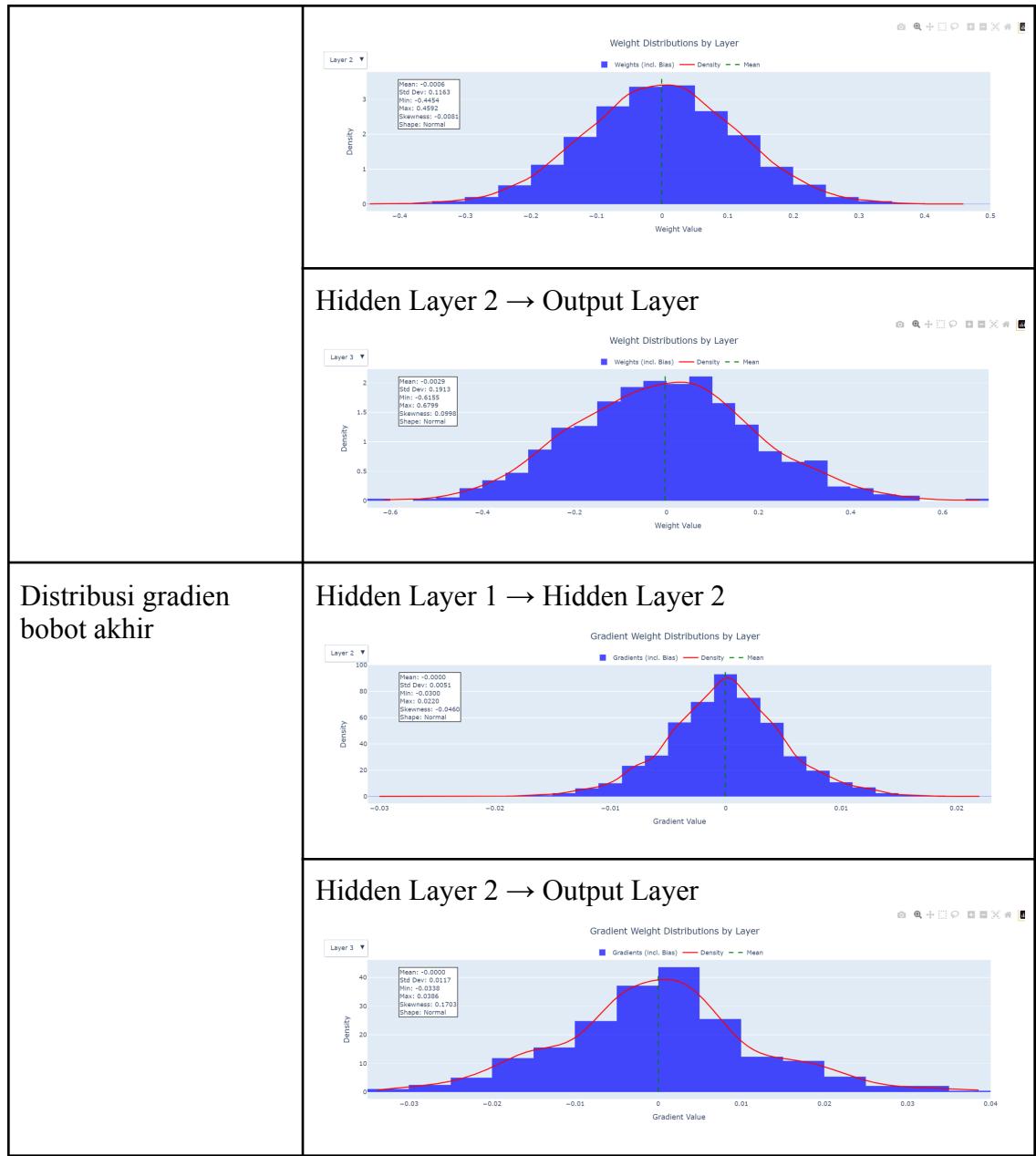
	precision	recall	f1-score	support
0	0.94	0.97	0.96	1402
1	0.97	0.97	0.97	1556
2	0.91	0.93	0.92	1368
3	0.89	0.89	0.89	1339
4	0.92	0.92	0.92	1395
5	0.92	0.87	0.89	1261
6	0.94	0.93	0.94	1418
7	0.94	0.90	0.92	1463
8	0.88	0.90	0.89	1362
9	0.88	0.90	0.89	1436
accuracy			0.92	14000
macro avg	0.92	0.92	0.92	14000
weighted avg	0.92	0.92	0.92	14000

Grafik loss



Distribusi bobot akhir

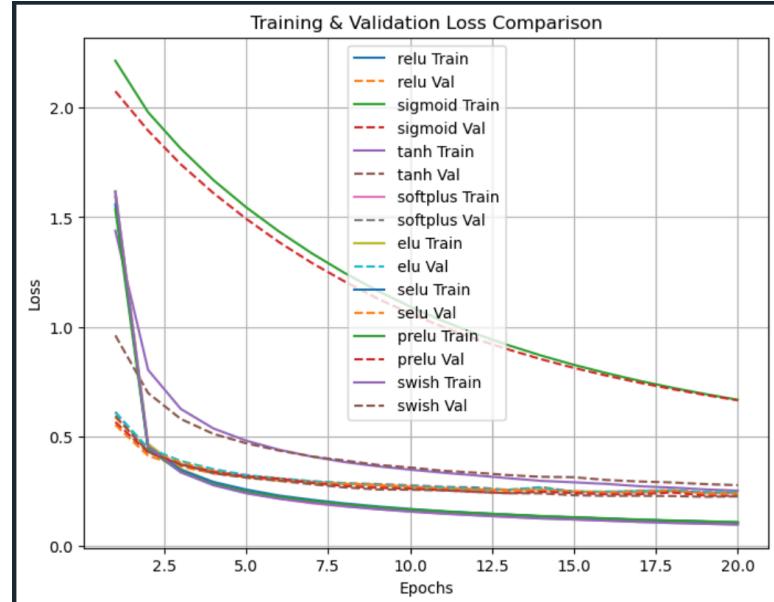
Hidden Layer 1 → Hidden Layer 2



2.2.2.3 Fungsi Aktivasi swish

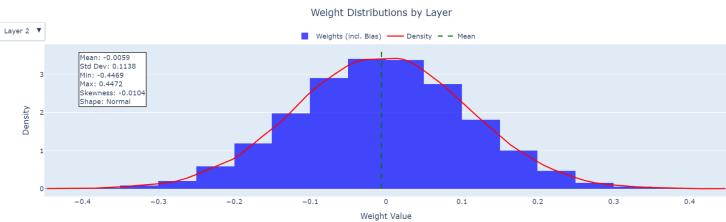
<p>Hasil akhir prediksi</p>	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="11">Confusion Matrix</th> </tr> <tr> <th></th> <th>0</th> <th>1</th> <th>2</th> <th>4</th> <th>1</th> <th>8</th> <th>5</th> <th>5</th> <th>8</th> <th>1</th> </tr> </thead> <tbody> <tr> <th>0</th> <td>1368</td> <td>0</td> <td>2</td> <td>4</td> <td>1</td> <td>8</td> <td>5</td> <td>5</td> <td>8</td> <td>1</td> </tr> <tr> <th>1</th> <td>1</td> <td>1530</td> <td>9</td> <td>3</td> <td>4</td> <td>0</td> <td>1</td> <td>0</td> <td>5</td> <td>3</td> </tr> <tr> <th>2</th> <td>11</td> <td>7</td> <td>1314</td> <td>6</td> <td>2</td> <td>5</td> <td>4</td> <td>8</td> <td>8</td> <td>3</td> </tr> <tr> <th>3</th> <td>1</td> <td>3</td> <td>28</td> <td>1223</td> <td>0</td> <td>20</td> <td>2</td> <td>11</td> <td>34</td> <td>17</td> </tr> <tr> <th>4</th> <td>2</td> <td>5</td> <td>11</td> <td>0</td> <td>1317</td> <td>2</td> <td>8</td> <td>8</td> <td>8</td> <td>34</td> </tr> <tr> <th>5</th> <td>8</td> <td>2</td> <td>7</td> <td>26</td> <td>5</td> <td>1171</td> <td>11</td> <td>3</td> <td>17</td> <td>11</td> </tr> <tr> <th>6</th> <td>16</td> <td>2</td> <td>14</td> <td>1</td> <td>10</td> <td>13</td> <td>1358</td> <td>0</td> <td>3</td> <td>1</td> </tr> <tr> <th>7</th> <td>6</td> <td>10</td> <td>24</td> <td>4</td> <td>11</td> <td>1</td> <td>1</td> <td>1367</td> <td>4</td> <td>35</td> </tr> <tr> <th>8</th> <td>2</td> <td>14</td> <td>13</td> <td>14</td> <td>8</td> <td>16</td> <td>3</td> <td>1</td> <td>1278</td> <td>13</td> </tr> <tr> <th>9</th> <td>13</td> <td>6</td> <td>2</td> <td>14</td> <td>45</td> <td>9</td> <td>1</td> <td>22</td> <td>11</td> <td>1313</td> </tr> <tr> <th></th> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> <tr> <th></th> <th>Predicted Labels</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> </tr> </tbody> </table>	Confusion Matrix												0	1	2	4	1	8	5	5	8	1	0	1368	0	2	4	1	8	5	5	8	1	1	1	1530	9	3	4	0	1	0	5	3	2	11	7	1314	6	2	5	4	8	8	3	3	1	3	28	1223	0	20	2	11	34	17	4	2	5	11	0	1317	2	8	8	8	34	5	8	2	7	26	5	1171	11	3	17	11	6	16	2	14	1	10	13	1358	0	3	1	7	6	10	24	4	11	1	1	1367	4	35	8	2	14	13	14	8	16	3	1	1278	13	9	13	6	2	14	45	9	1	22	11	1313		0	1	2	3	4	5	6	7	8	9		Predicted Labels									
Confusion Matrix																																																																																																																																																											
	0	1	2	4	1	8	5	5	8	1																																																																																																																																																	
0	1368	0	2	4	1	8	5	5	8	1																																																																																																																																																	
1	1	1530	9	3	4	0	1	0	5	3																																																																																																																																																	
2	11	7	1314	6	2	5	4	8	8	3																																																																																																																																																	
3	1	3	28	1223	0	20	2	11	34	17																																																																																																																																																	
4	2	5	11	0	1317	2	8	8	8	34																																																																																																																																																	
5	8	2	7	26	5	1171	11	3	17	11																																																																																																																																																	
6	16	2	14	1	10	13	1358	0	3	1																																																																																																																																																	
7	6	10	24	4	11	1	1	1367	4	35																																																																																																																																																	
8	2	14	13	14	8	16	3	1	1278	13																																																																																																																																																	
9	13	6	2	14	45	9	1	22	11	1313																																																																																																																																																	
	0	1	2	3	4	5	6	7	8	9																																																																																																																																																	
	Predicted Labels																																																																																																																																																										
<p>Skor validasi</p>	<pre> Weighted F1-Score: 0.9456 Macro F1-Score: 0.9451 Cohen's Kappa: 0.9396 ROC-AUC (Weighted): 0.9967 Validation Log Loss: 0.2271 Classification Report: precision recall f1-score support 0 0.96 0.98 0.97 1402 1 0.97 0.98 0.98 1556 2 0.92 0.96 0.94 1368 3 0.94 0.91 0.93 1339 4 0.94 0.94 0.94 1395 5 0.94 0.93 0.93 1261 6 0.97 0.96 0.97 1418 7 0.96 0.93 0.95 1463 8 0.93 0.94 0.93 1362 9 0.92 0.91 0.92 1436 accuracy 0.95 14000 macro avg 0.95 0.95 0.95 14000 weighted avg 0.95 0.95 0.95 14000 </pre>																																																																																																																																																										

Grafik loss



Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

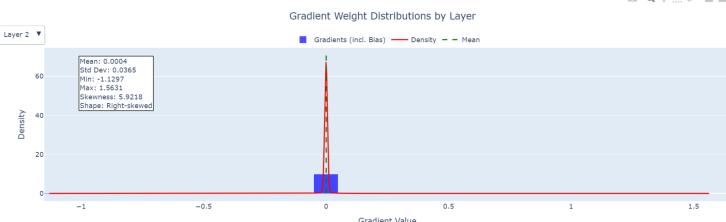


Hidden Layer 2 → Output Layer



Distribusi gradien bobot akhir

Hidden Layer 1 → Hidden Layer 2



Hidden Layer 2 → Output Layer



Diketahui bahwa sigmoid memiliki loss paling besar dimana ini menyebabkan akurasi sangat kecil. Swish memiliki loss yang kecil sehingga akurasinya besar.

2.2.4. Pengaruh *Learning Rate*

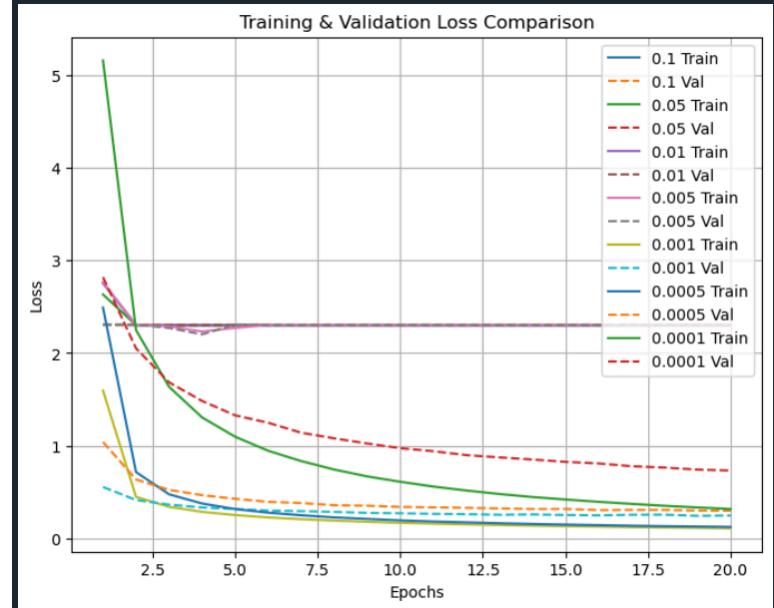
Parameter kontrol yang digunakan untuk membandingkan pengaruh antara learning rate yang digunakan adalah :

Jumlah layer	: 4 (input layer, hidden layer 1, hidden layer 2, output layer)
Neuron tiap layer	: 784 neuron, 150 neuron, 75 neuron, 10 neuron
Fungsi aktivasi	: ReLU (hidden layer 1), ReLU (hidden layer 2), softmax
(output layer)	
Inisialisasi Bobot	: Normal He Initialization
Fungsi loss	: Categorical Cross-Entropy
Batch size	: 64
Epoch maksimal	: 20

2.2.3.1 Learning Rate 0.1

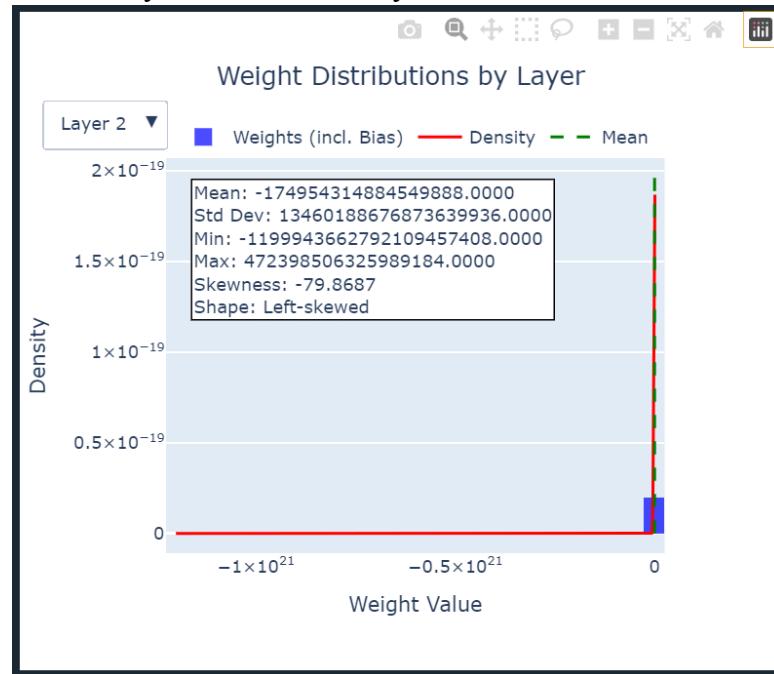
Hasil akhir prediksi	<p>Confusion Matrix</p> <table border="1"> <thead> <tr> <th></th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th></tr> </thead> <tbody> <tr> <th>0</th><td>1402</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>1</th><td>0</td><td>1556</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>2</th><td>0</td><td>1368</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>3</th><td>0</td><td>1339</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>4</th><td>0</td><td>1393</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>5</th><td>0</td><td>1261</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>6</th><td>0</td><td>1418</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>7</th><td>0</td><td>1463</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>8</th><td>0</td><td>1362</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <th>9</th><td>0</td><td>1436</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </tbody> </table>		0	1	2	3	4	5	6	7	8	9	0	1402	0	0	0	0	0	0	0	0	0	1	0	1556	0	0	0	0	0	0	0	0	2	0	1368	0	0	0	0	0	0	0	0	3	0	1339	0	0	0	0	0	0	0	0	4	0	1393	2	0	0	0	0	0	0	0	5	0	1261	0	0	0	0	0	0	0	0	6	0	1418	0	0	0	0	0	0	0	0	7	0	1463	0	0	0	0	0	0	0	0	8	0	1362	0	0	0	0	0	0	0	0	9	0	1436	0	0	0	0	0	0	0	0
	0	1	2	3	4	5	6	7	8	9																																																																																																																
0	1402	0	0	0	0	0	0	0	0	0																																																																																																																
1	0	1556	0	0	0	0	0	0	0	0																																																																																																																
2	0	1368	0	0	0	0	0	0	0	0																																																																																																																
3	0	1339	0	0	0	0	0	0	0	0																																																																																																																
4	0	1393	2	0	0	0	0	0	0	0																																																																																																																
5	0	1261	0	0	0	0	0	0	0	0																																																																																																																
6	0	1418	0	0	0	0	0	0	0	0																																																																																																																
7	0	1463	0	0	0	0	0	0	0	0																																																																																																																
8	0	1362	0	0	0	0	0	0	0	0																																																																																																																
9	0	1436	0	0	0	0	0	0	0	0																																																																																																																
Skor validasi	<pre> Weighted F1-Score: 0.0222 Macro F1-Score: 0.0200 Cohen's Kappa: 0.0000 ROC-AUC (Weighted): 0.5000 Validation Log Loss: 2.3067 Classification Report: precision recall f1-score support 0 0.00 0.00 0.00 1402 1 0.11 1.00 0.20 1556 2 0.00 0.00 0.00 1368 3 0.00 0.00 0.00 1339 4 0.00 0.00 0.00 1395 5 0.00 0.00 0.00 1261 6 0.00 0.00 0.00 1418 7 0.00 0.00 0.00 1463 8 0.00 0.00 0.00 1362 9 0.00 0.00 0.00 1436 accuracy 0.11 14000 macro avg 0.01 0.10 0.02 14000 weighted avg 0.01 0.11 0.02 14000 </pre>																																																																																																																									

Grafik loss

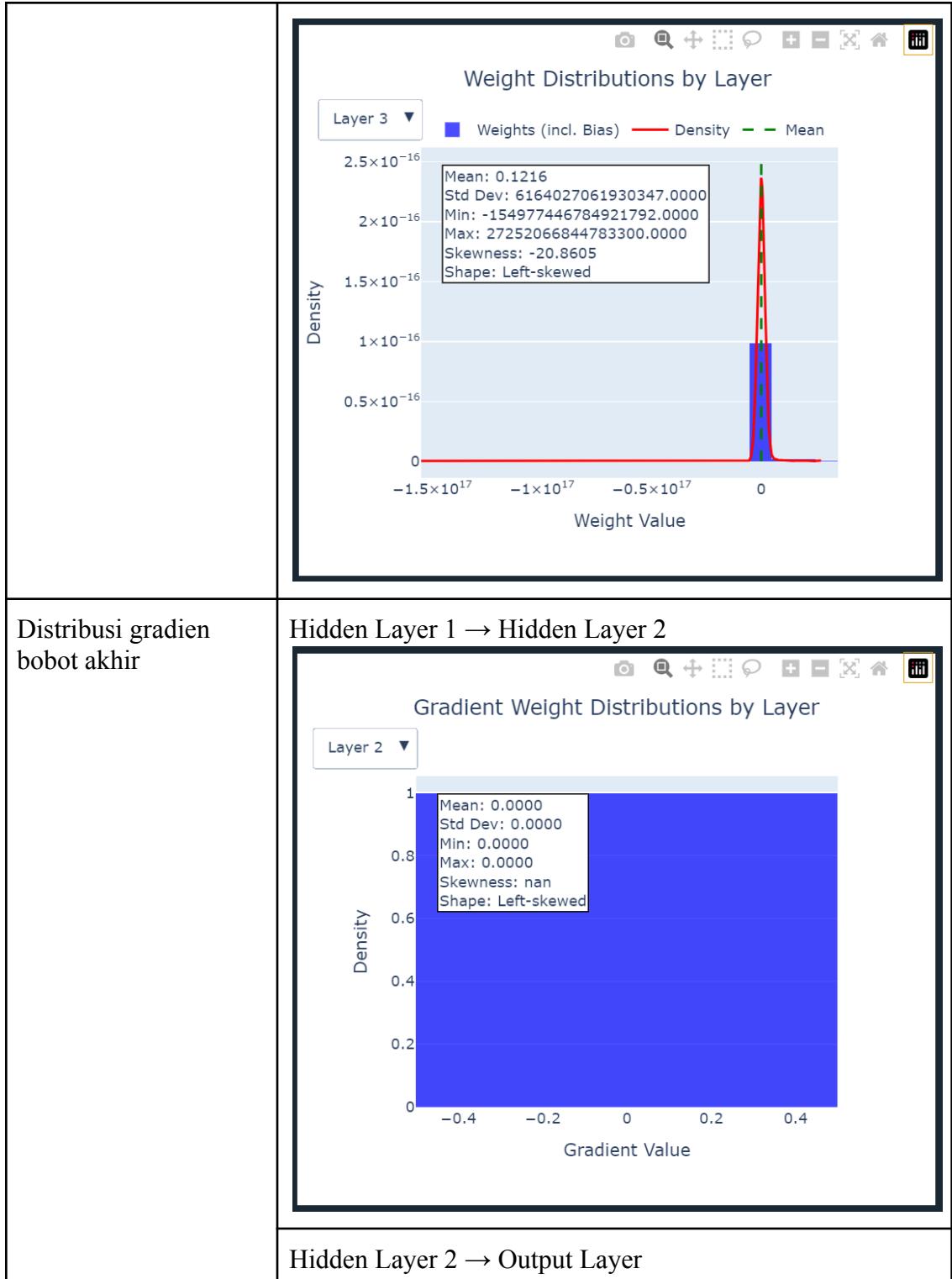


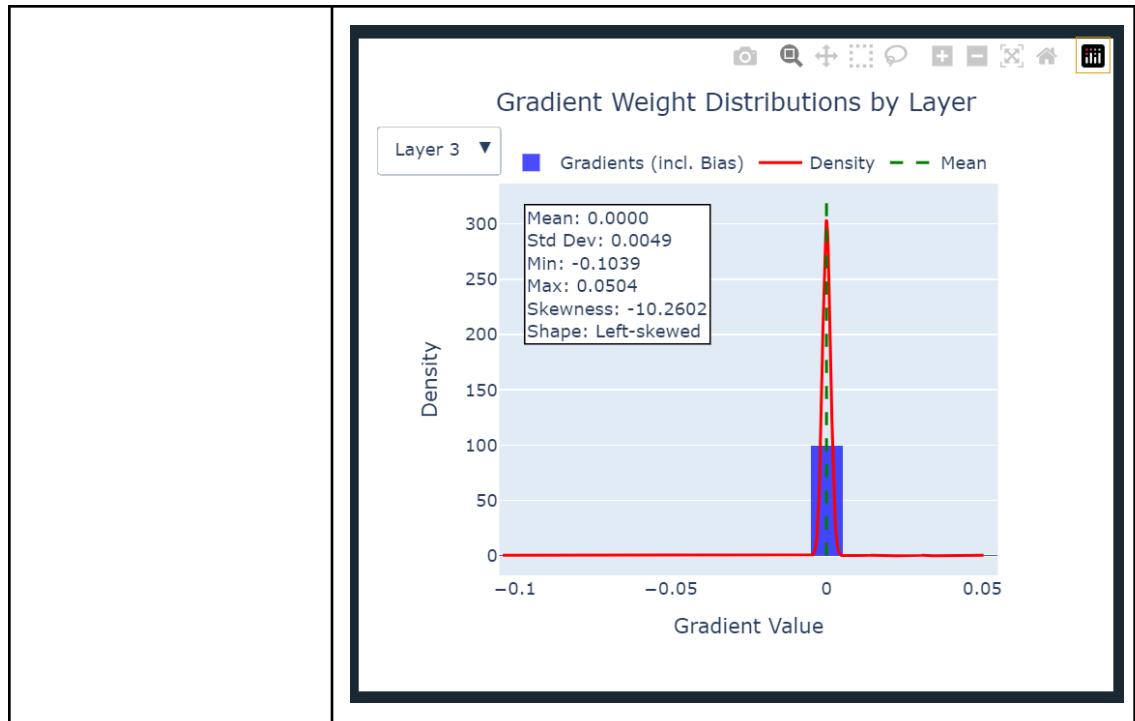
Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

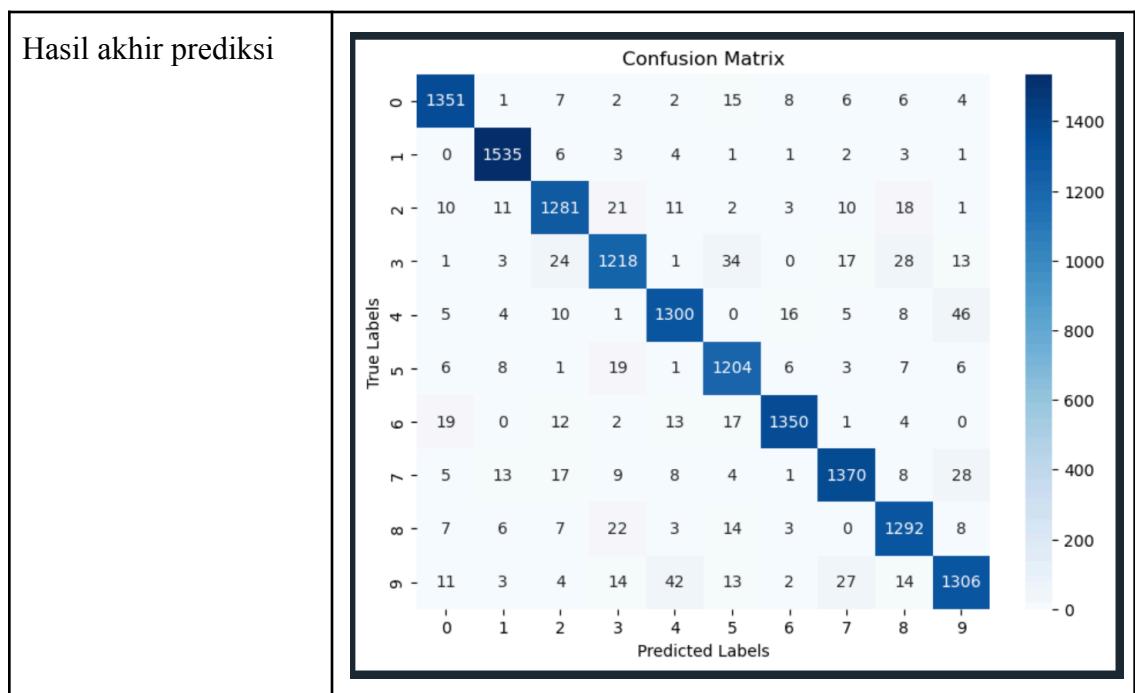


Hidden Layer 2 → Output Layer





2.2.3.2 Learning Rate 0.001



Skor validasi

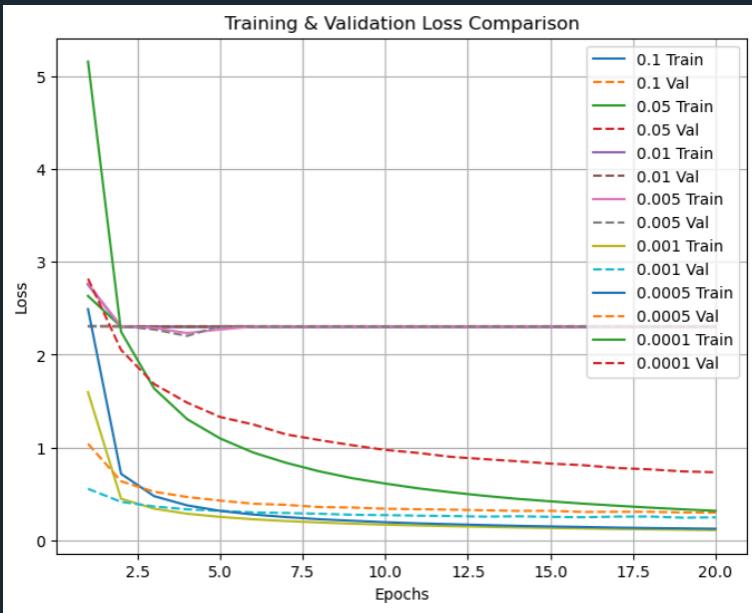
```
Weighted F1-Score: 0.9433  
Macro F1-Score: 0.9428  
Cohen's Kappa: 0.9370  
ROC-AUC (Weighted): 0.9961
```

```
Validation Log Loss: 0.2458
```

Classification Report:

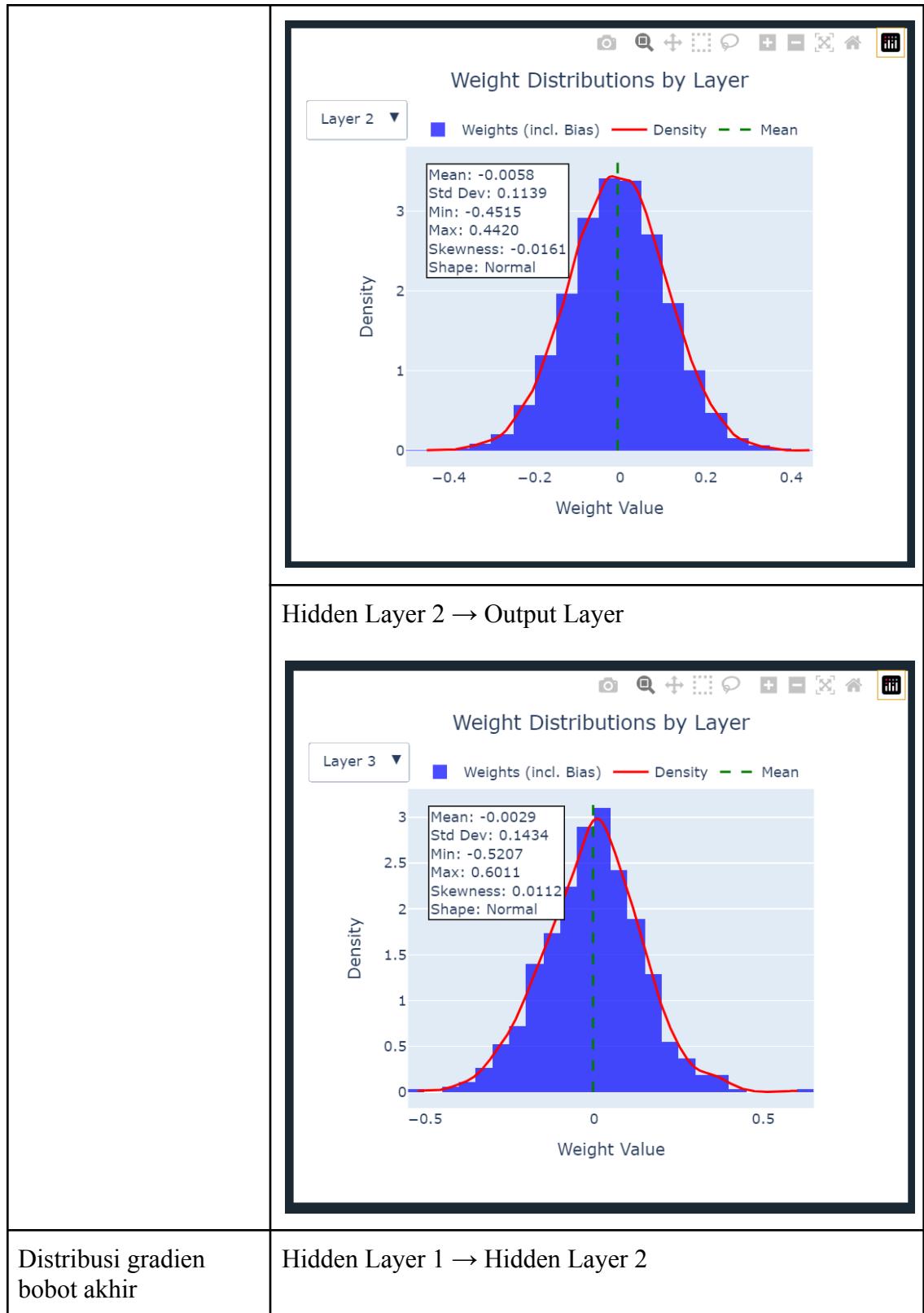
	precision	recall	f1-score	support
0	0.95	0.96	0.96	1402
1	0.97	0.99	0.98	1556
2	0.94	0.94	0.94	1368
3	0.93	0.91	0.92	1339
4	0.94	0.93	0.94	1395
5	0.92	0.95	0.94	1261
6	0.97	0.95	0.96	1418
7	0.95	0.94	0.94	1463
8	0.93	0.95	0.94	1362
9	0.92	0.91	0.92	1436
accuracy			0.94	14000
macro avg	0.94	0.94	0.94	14000
weighted avg	0.94	0.94	0.94	14000

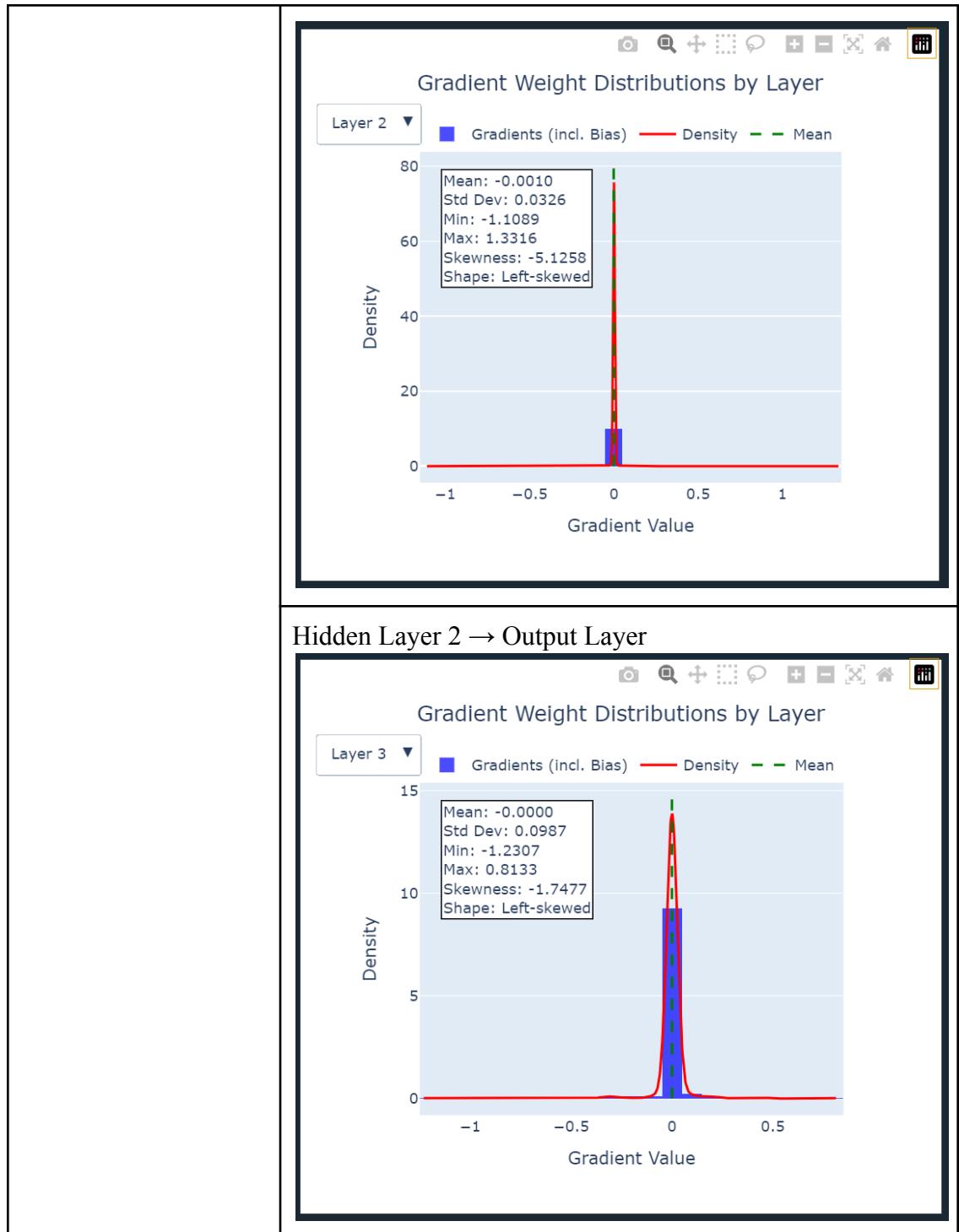
Grafik loss



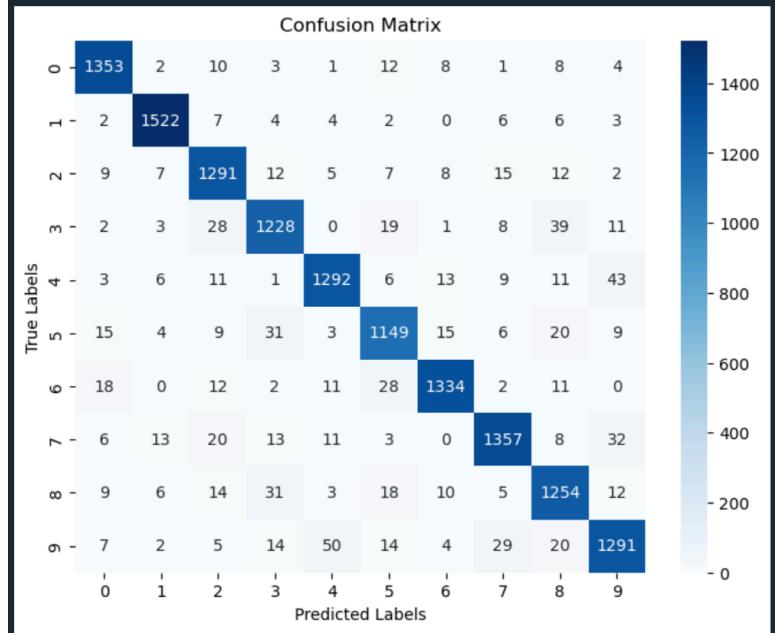
Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2

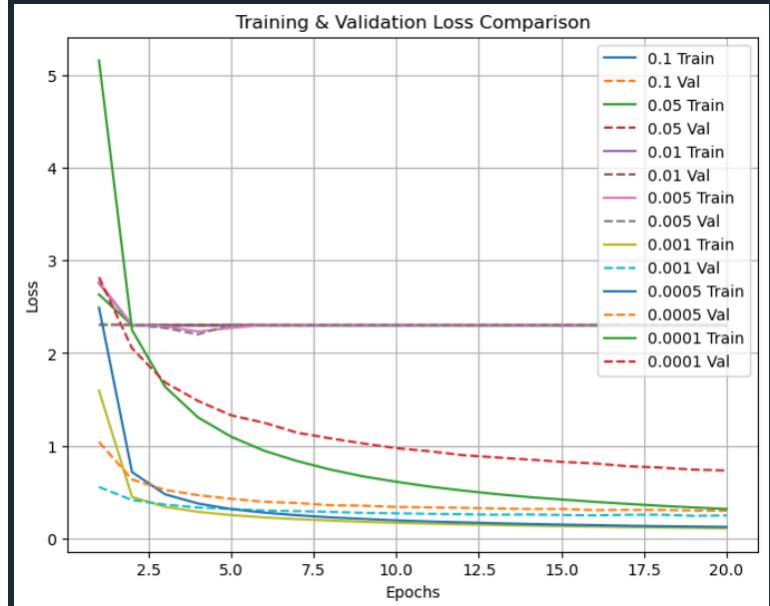




2.2.3.3 Learning Rate 0.005

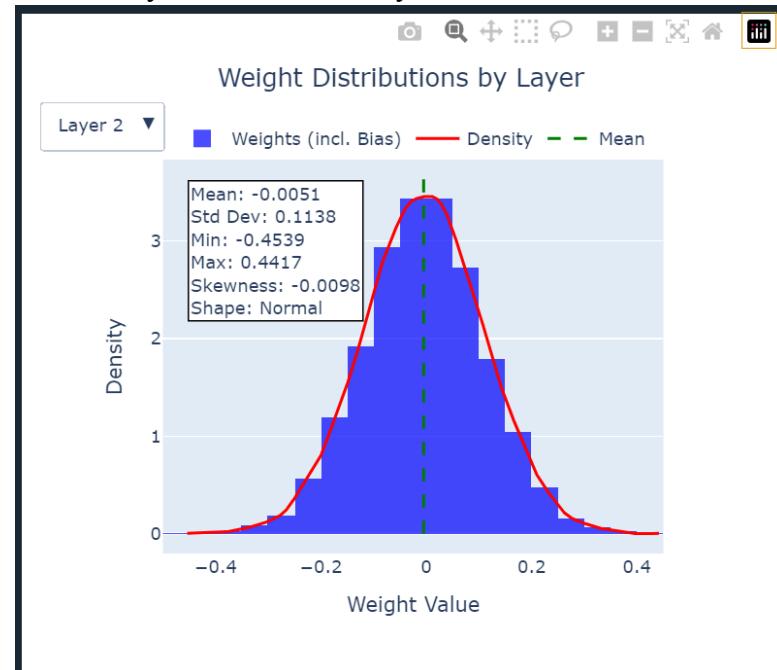
Hasil akhir prediksi	 <table border="1" style="margin-top: 10px; border-collapse: collapse; width: 100%; text-align: center;"> <thead> <tr> <th colspan="10">Confusion Matrix</th> </tr> <tr> <th></th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> </tr> </thead> <tbody> <tr> <th>0</th> <td>1353</td> <td>2</td> <td>10</td> <td>3</td> <td>1</td> <td>12</td> <td>8</td> <td>1</td> <td>8</td> <td>4</td> </tr> <tr> <th>1</th> <td>2</td> <td>1522</td> <td>7</td> <td>4</td> <td>4</td> <td>2</td> <td>0</td> <td>6</td> <td>6</td> <td>3</td> </tr> <tr> <th>2</th> <td>9</td> <td>7</td> <td>1291</td> <td>12</td> <td>5</td> <td>7</td> <td>8</td> <td>15</td> <td>12</td> <td>2</td> </tr> <tr> <th>3</th> <td>2</td> <td>3</td> <td>28</td> <td>1228</td> <td>0</td> <td>19</td> <td>1</td> <td>8</td> <td>39</td> <td>11</td> </tr> <tr> <th>4</th> <td>3</td> <td>6</td> <td>11</td> <td>1</td> <td>1292</td> <td>6</td> <td>13</td> <td>9</td> <td>11</td> <td>43</td> </tr> <tr> <th>5</th> <td>15</td> <td>4</td> <td>9</td> <td>31</td> <td>3</td> <td>1149</td> <td>15</td> <td>6</td> <td>20</td> <td>9</td> </tr> <tr> <th>6</th> <td>18</td> <td>0</td> <td>12</td> <td>2</td> <td>11</td> <td>28</td> <td>1334</td> <td>2</td> <td>11</td> <td>0</td> </tr> <tr> <th>7</th> <td>6</td> <td>13</td> <td>20</td> <td>13</td> <td>11</td> <td>3</td> <td>0</td> <td>1357</td> <td>8</td> <td>32</td> </tr> <tr> <th>8</th> <td>9</td> <td>6</td> <td>14</td> <td>31</td> <td>3</td> <td>18</td> <td>10</td> <td>5</td> <td>1254</td> <td>12</td> </tr> <tr> <th>9</th> <td>7</td> <td>2</td> <td>5</td> <td>14</td> <td>50</td> <td>14</td> <td>4</td> <td>29</td> <td>20</td> <td>1291</td> </tr> <tr> <th></th> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> <tr> <th>Predicted Labels</th> <td></td> </tr> </tbody> </table>	Confusion Matrix											0	1	2	3	4	5	6	7	8	9	0	1353	2	10	3	1	12	8	1	8	4	1	2	1522	7	4	4	2	0	6	6	3	2	9	7	1291	12	5	7	8	15	12	2	3	2	3	28	1228	0	19	1	8	39	11	4	3	6	11	1	1292	6	13	9	11	43	5	15	4	9	31	3	1149	15	6	20	9	6	18	0	12	2	11	28	1334	2	11	0	7	6	13	20	13	11	3	0	1357	8	32	8	9	6	14	31	3	18	10	5	1254	12	9	7	2	5	14	50	14	4	29	20	1291		0	1	2	3	4	5	6	7	8	9	Predicted Labels										
Confusion Matrix																																																																																																																																																										
	0	1	2	3	4	5	6	7	8	9																																																																																																																																																
0	1353	2	10	3	1	12	8	1	8	4																																																																																																																																																
1	2	1522	7	4	4	2	0	6	6	3																																																																																																																																																
2	9	7	1291	12	5	7	8	15	12	2																																																																																																																																																
3	2	3	28	1228	0	19	1	8	39	11																																																																																																																																																
4	3	6	11	1	1292	6	13	9	11	43																																																																																																																																																
5	15	4	9	31	3	1149	15	6	20	9																																																																																																																																																
6	18	0	12	2	11	28	1334	2	11	0																																																																																																																																																
7	6	13	20	13	11	3	0	1357	8	32																																																																																																																																																
8	9	6	14	31	3	18	10	5	1254	12																																																																																																																																																
9	7	2	5	14	50	14	4	29	20	1291																																																																																																																																																
	0	1	2	3	4	5	6	7	8	9																																																																																																																																																
Predicted Labels																																																																																																																																																										
Skor validasi	<pre> Weighted F1-Score: 0.9336 Macro F1-Score: 0.9328 Cohen's Kappa: 0.9262 ROC-AUC (Weighted): 0.9952 Validation Log Loss: 0.2979 Classification Report: precision recall f1-score support 0 0.95 0.97 0.96 1402 1 0.97 0.98 0.98 1556 2 0.92 0.94 0.93 1368 3 0.92 0.92 0.92 1339 4 0.94 0.93 0.93 1395 5 0.91 0.91 0.91 1261 6 0.96 0.94 0.95 1418 7 0.94 0.93 0.94 1463 8 0.90 0.92 0.91 1362 9 0.92 0.90 0.91 1436 accuracy 0.93 14000 macro avg 0.93 0.93 0.93 14000 weighted avg 0.93 0.93 0.93 14000 </pre>																																																																																																																																																									

Grafik loss

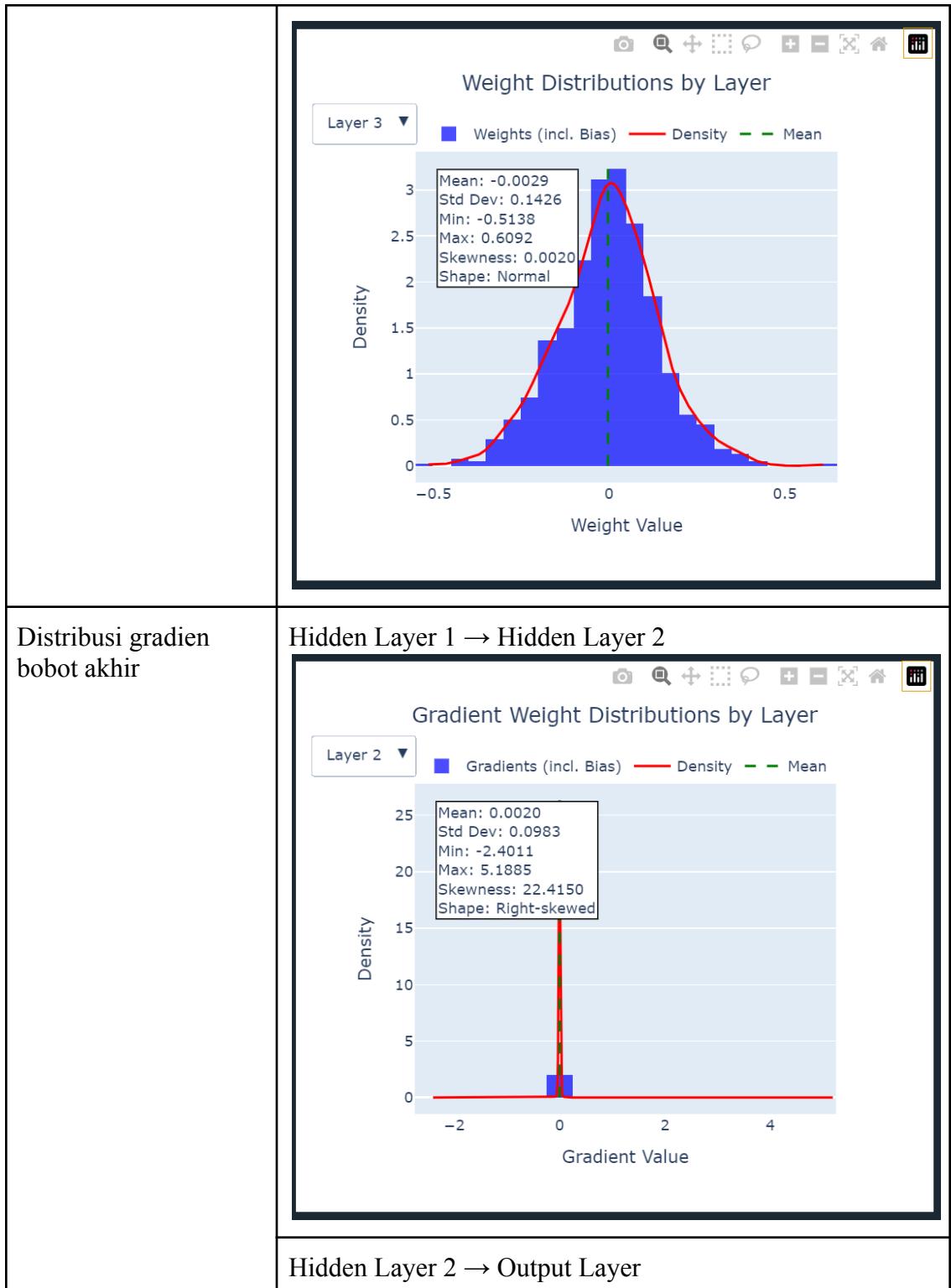


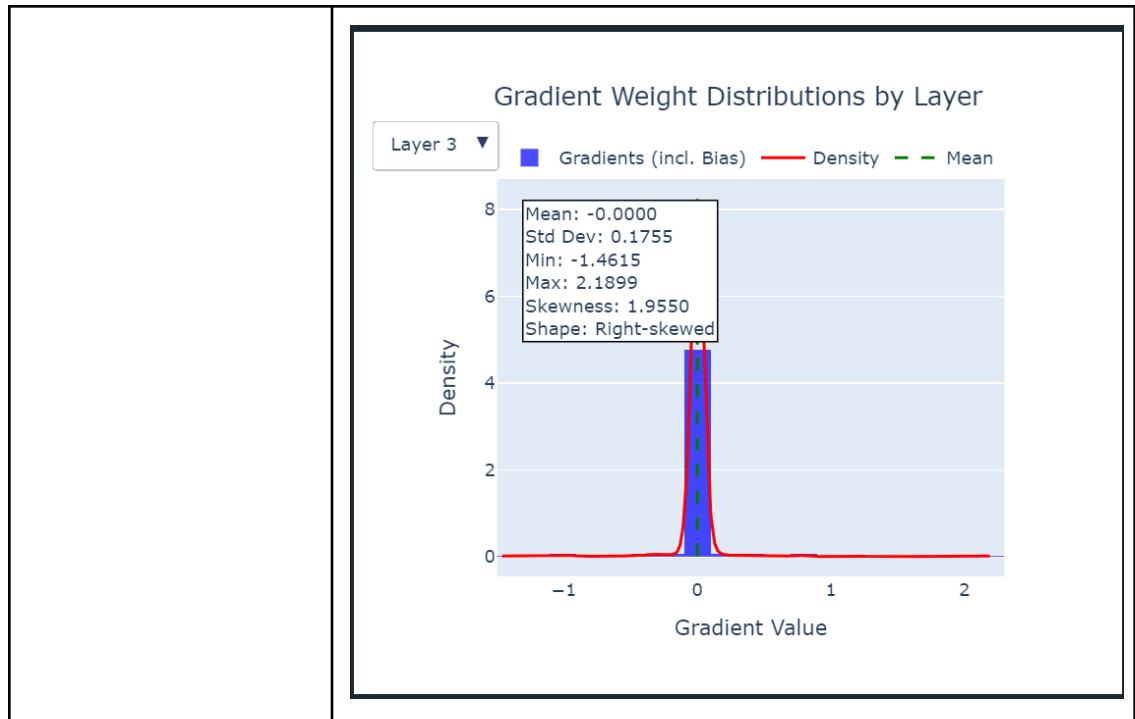
Distribusi bobot akhir

Hidden Layer 1 → Hidden Layer 2



Hidden Layer 2 → Output Layer





Setelah dilakukan beberapa percobaan dengan mengganti learning rate, *learning rate* terbaik berada di angka 0.001. Jika learning rate lebih kecil dari angka tersebut, loss yang dihasilkan model akan semakin memburuk sebanding dengan semakin mengecilnya *learning rate*.

2.2.5. Pengaruh Inisialisasi Bobot

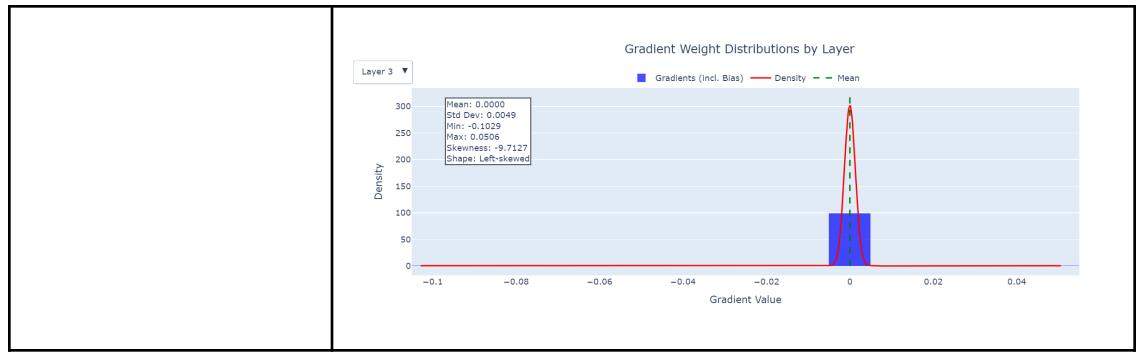
Parameter kontrol yang digunakan untuk membandingkan pengaruh antara metode inisialisasi bobot yang diimplementasikan adalah :

Jumlah layer	:	4 (input layer, hidden layer 1, hidden layer 2, output layer)
Neuron tiap layer	:	784 neuron, 150 neuron, 75 neuron, 10 neuron
Fungsi aktivasi	:	ReLU (hidden layer 1), ReLU (hidden layer 2), softmax (output layer)
Fungsi loss	:	Categorical Cross-Entropy
Batch size	:	64
Learning rate	:	0.001
Epoch maksimal	:	20

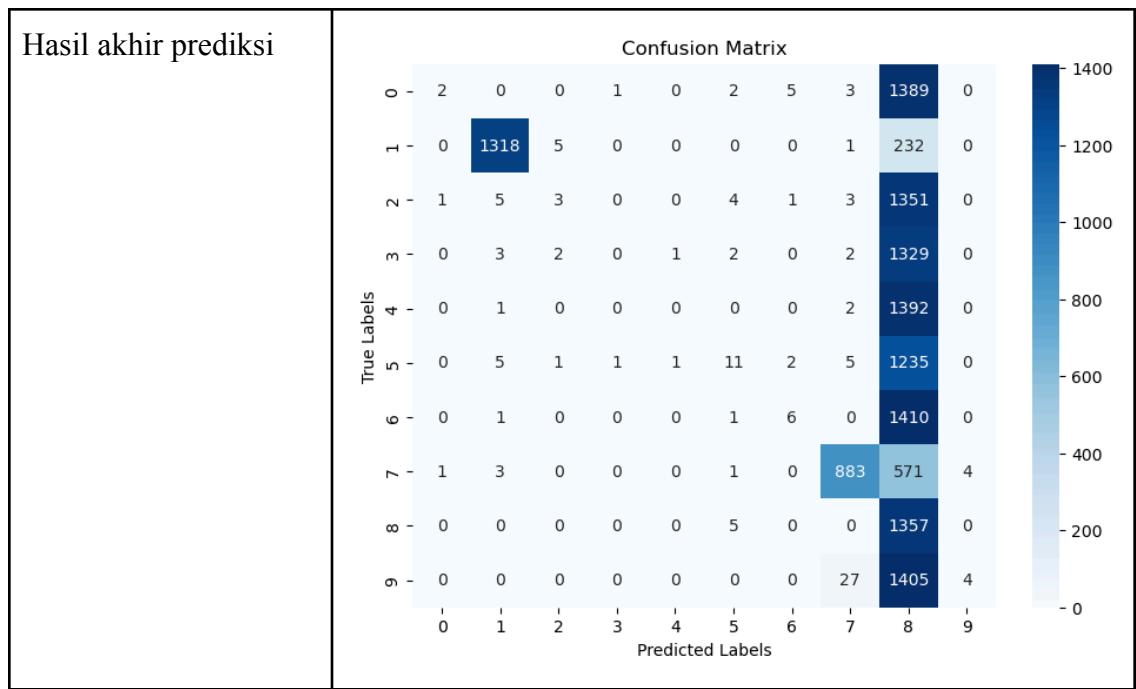
2.2.4.1. Zero Initialization

Hasil akhir prediksi	<p style="text-align: center;">Confusion Matrix</p> <table border="1" style="margin-top: 10px; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2"></th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> </tr> <tr> <th rowspan="2">True Labels</th> <th>Predicted Labels</th> <th>0</th> <td>1402</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> </thead> <tbody> <tr> <th>0</th> <td>1402</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>1</th> <td>1556</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>2</th> <td>1368</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>3</th> <td>1339</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>4</th> <td>1395</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>5</th> <td>1261</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>6</th> <td>1418</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>7</th> <td>1463</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>8</th> <td>1362</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> <tr> <th>9</th> <td>1436</td> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> <th>0</th> </tr> </tbody> </table>			0	1	2	3	4	5	6	7	8	9	True Labels	Predicted Labels	0	1402	0	0	0	0	0	0	0	0	0	1402	0	0	0	0	0	0	0	0	0	1	1556	0	0	0	0	0	0	0	0	0	2	1368	0	0	0	0	0	0	0	0	0	3	1339	0	0	0	0	0	0	0	0	0	4	1395	0	0	0	0	0	0	0	0	0	5	1261	0	0	0	0	0	0	0	0	0	6	1418	0	0	0	0	0	0	0	0	0	7	1463	0	0	0	0	0	0	0	0	0	8	1362	0	0	0	0	0	0	0	0	0	9	1436	0	0	0	0	0	0	0	0	0
		0	1	2	3	4	5	6	7	8	9																																																																																																																												
True Labels	Predicted Labels	0	1402	0	0	0	0	0	0	0	0																																																																																																																												
	0	1402	0	0	0	0	0	0	0	0	0																																																																																																																												
1	1556	0	0	0	0	0	0	0	0	0																																																																																																																													
2	1368	0	0	0	0	0	0	0	0	0																																																																																																																													
3	1339	0	0	0	0	0	0	0	0	0																																																																																																																													
4	1395	0	0	0	0	0	0	0	0	0																																																																																																																													
5	1261	0	0	0	0	0	0	0	0	0																																																																																																																													
6	1418	0	0	0	0	0	0	0	0	0																																																																																																																													
7	1463	0	0	0	0	0	0	0	0	0																																																																																																																													
8	1362	0	0	0	0	0	0	0	0	0																																																																																																																													
9	1436	0	0	0	0	0	0	0	0	0																																																																																																																													
Skor validasi	<pre> Weighted F1-Score: 0.0222 Macro F1-Score: 0.0200 Cohen's Kappa: 0.0000 ROC-AUC (Weighted): 0.5000 Validation Log Loss: 2.3017 Classification Report: precision recall f1-score support 0 0.00 0.00 0.00 1402 1 0.11 1.00 0.20 1556 2 0.00 0.00 0.00 1368 3 0.00 0.00 0.00 1339 4 0.00 0.00 0.00 1395 5 0.00 0.00 0.00 1261 6 0.00 0.00 0.00 1418 7 0.00 0.00 0.00 1463 8 0.00 0.00 0.00 1362 9 0.00 0.00 0.00 1436 accuracy 0.11 14000 macro avg 0.01 0.10 0.02 14000 weighted avg 0.01 0.11 0.02 14000 </pre>																																																																																																																																						

Grafik loss	<p>Training Loss</p> <p>Models — Training Loss — Validation Loss</p> <p>Epoch</p>
Distribusi bobot akhir	<p>Input Layer → Hidden Layer 1 Distribusi bobot tidak dapat digambarkan akibat perbedaan bobot yang sangat kecil pada seluruh node yang terlibat.</p>
	<p>Hidden Layer 1 → Hidden Layer 2 Distribusi bobot tidak dapat digambarkan akibat perbedaan bobot yang sangat kecil pada seluruh node yang terlibat.</p>
	<p>Hidden Layer 2 → Output Layer Distribusi bobot tidak dapat digambarkan akibat perbedaan bobot yang sangat kecil pada seluruh node yang terlibat.</p>
Distribusi gradien bobot akhir	<p>Input Layer → Hidden Layer 1</p> <p>Gradient Weight Distributions by Layer</p> <p>Layer 1</p> <p>Density</p> <p>Gradient Value</p>
	<p>Hidden Layer 1 → Hidden Layer 2</p> <p>Gradient Weight Distributions by Layer</p> <p>Layer 2</p> <p>Density</p> <p>Gradient Value</p>
	<p>Hidden Layer 2 → Output Layer</p>



2.2.4.2. Random Uniform Initialization



Skor validasi

Weighted F1-Score: 0.2025
Macro F1-Score: 0.1897
Cohen's Kappa: 0.1742
ROC-AUC (Weighted): 0.6509

Validation Log Loss: 1.9861

Classification Report:

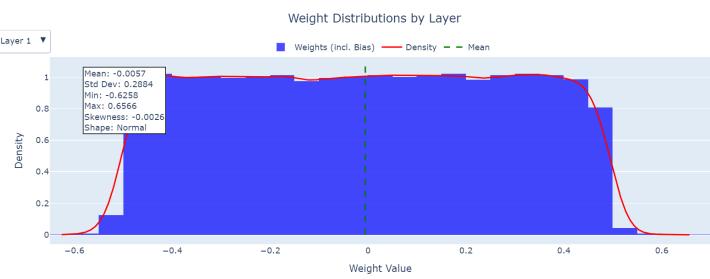
	precision	recall	f1-score	support
0	0.50	0.00	0.00	1402
1	0.99	0.85	0.91	1556
2	0.27	0.00	0.00	1368
3	0.00	0.00	0.00	1339
4	0.00	0.00	0.00	1395
5	0.42	0.01	0.02	1261
6	0.43	0.00	0.01	1418
7	0.95	0.60	0.74	1463
8	0.12	1.00	0.21	1362
9	0.50	0.00	0.01	1436
accuracy			0.26	14000
macro avg	0.42	0.25	0.19	14000
weighted avg	0.43	0.26	0.20	14000

Grafik loss

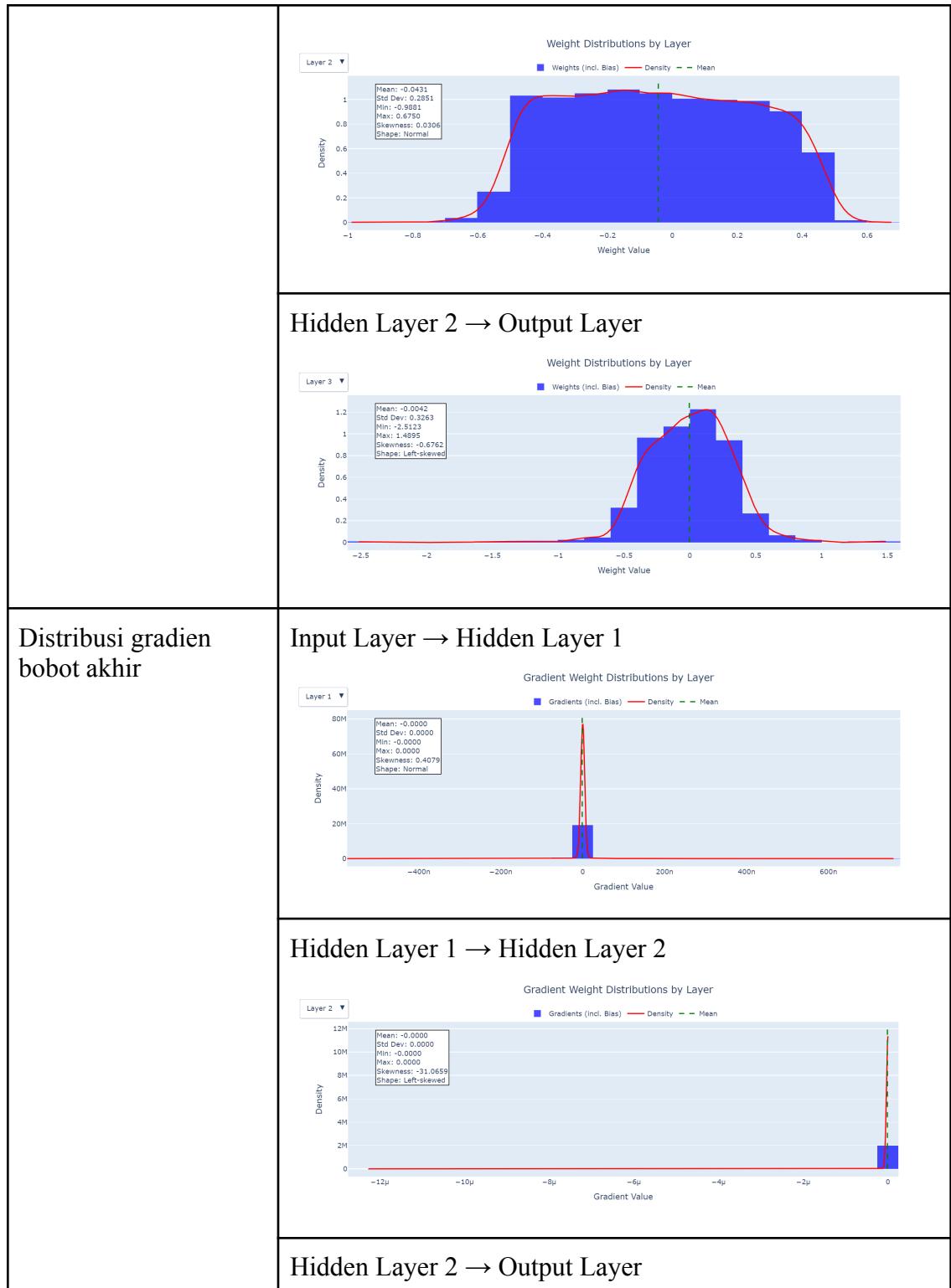


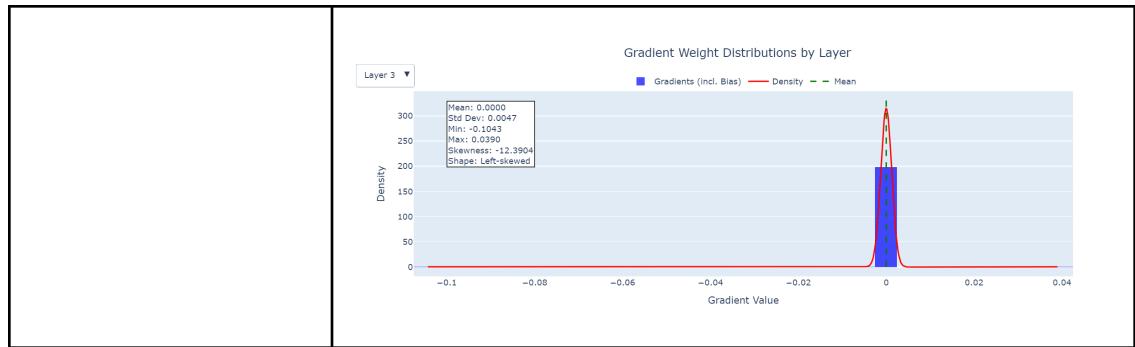
Distribusi bobot akhir

Input Layer → Hidden Layer 1

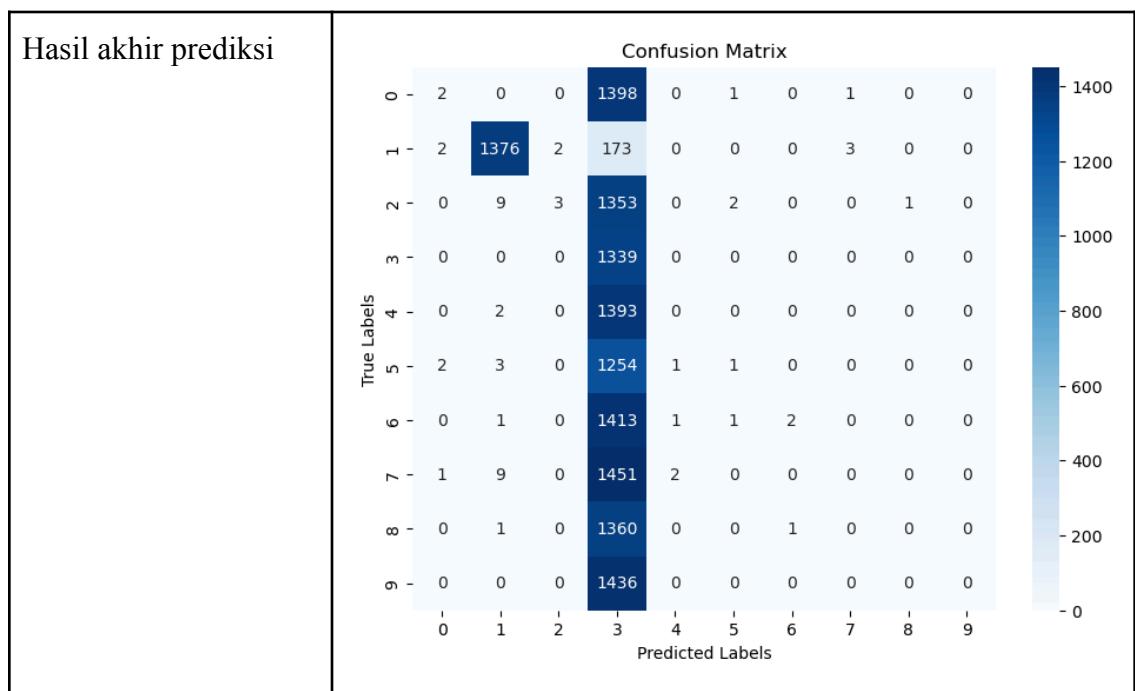


Hidden Layer 1 → Hidden Layer 2





2.2.4.3. Random Normal Initialization



Skor validasi

Weighted F1-Score: 0.1230
Macro F1-Score: 0.1135
Cohen's Kappa: 0.1078
ROC-AUC (Weighted): 0.5984

Validation Log Loss: 2.0882

Classification Report:

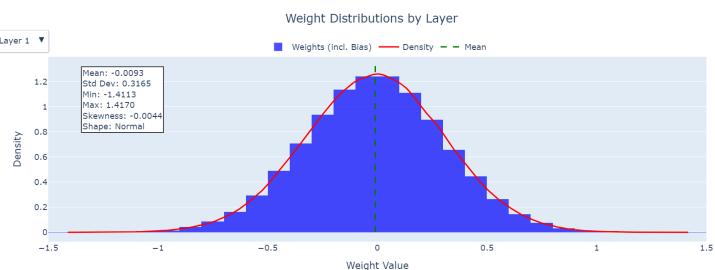
	precision	recall	f1-score	support
0	0.29	0.00	0.00	1402
1	0.98	0.88	0.93	1556
2	0.60	0.00	0.00	1368
3	0.11	1.00	0.19	1339
4	0.00	0.00	0.00	1395
5	0.20	0.00	0.00	1261
6	0.67	0.00	0.00	1418
7	0.00	0.00	0.00	1463
8	0.00	0.00	0.00	1362
9	0.00	0.00	0.00	1436
accuracy			0.19	14000
macro avg	0.28	0.19	0.11	14000
weighted avg	0.29	0.19	0.12	14000

Grafik loss

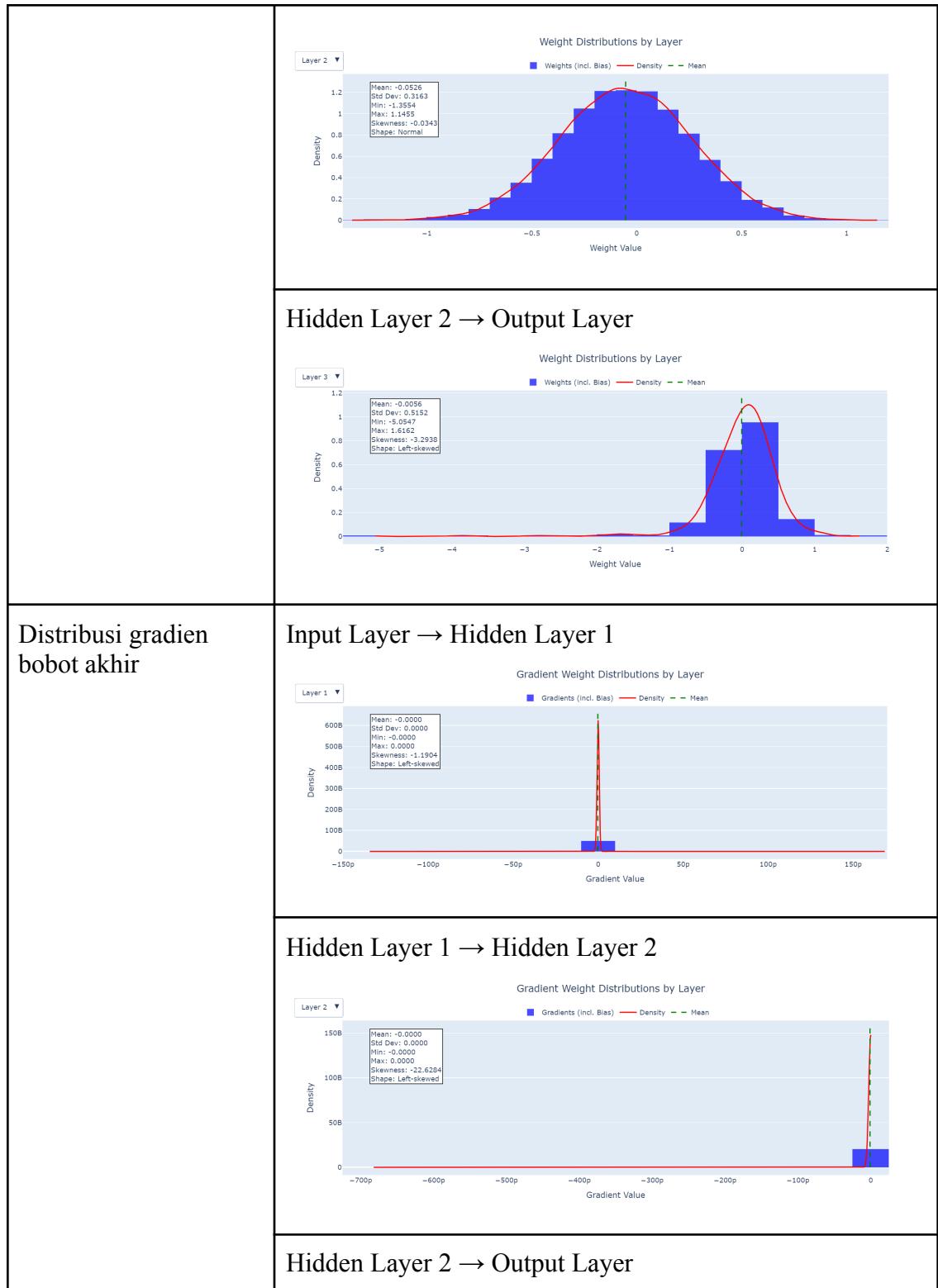


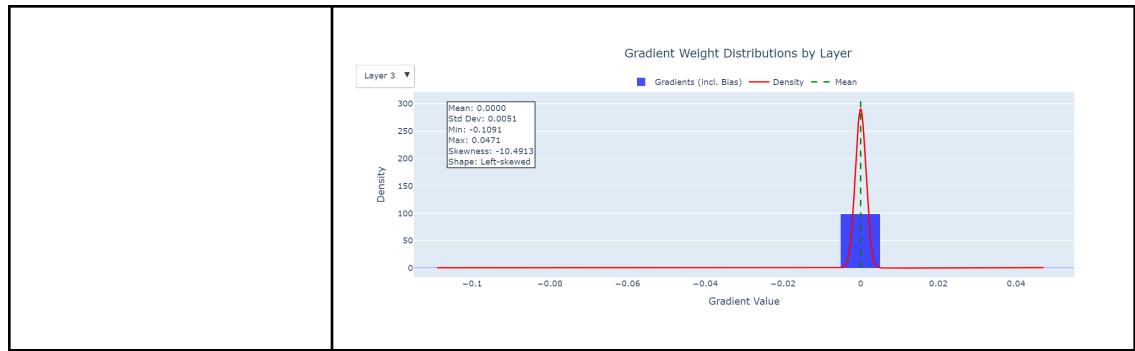
Distribusi bobot akhir

Input Layer → Hidden Layer 1

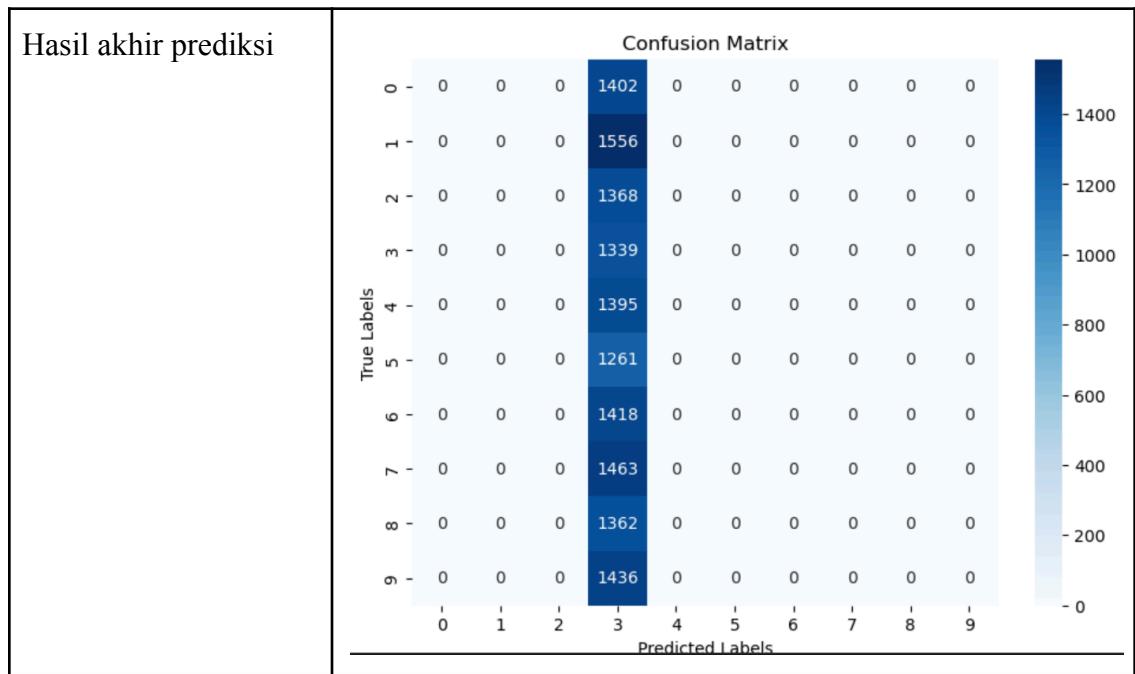


Hidden Layer 1 → Hidden Layer 2





2.2.4.4. Normal Xavier Initialization



Skor validasi

Weighted F1-Score: 0.0167
Macro F1-Score: 0.0175
Cohen's Kappa: 0.0000
ROC-AUC (Weighted): 0.5000

Validation Log Loss: 3.2067

Classification Report:

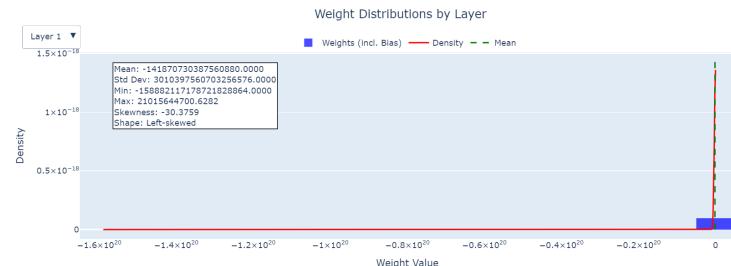
	precision	recall	f1-score	support
0	0.00	0.00	0.00	1402
1	0.00	0.00	0.00	1556
2	0.00	0.00	0.00	1368
3	0.10	1.00	0.17	1339
4	0.00	0.00	0.00	1395
5	0.00	0.00	0.00	1261
6	0.00	0.00	0.00	1418
7	0.00	0.00	0.00	1463
8	0.00	0.00	0.00	1362
9	0.00	0.00	0.00	1436
accuracy			0.10	14000
macro avg	0.01	0.10	0.02	14000
weighted avg	0.01	0.10	0.02	14000

Grafik loss

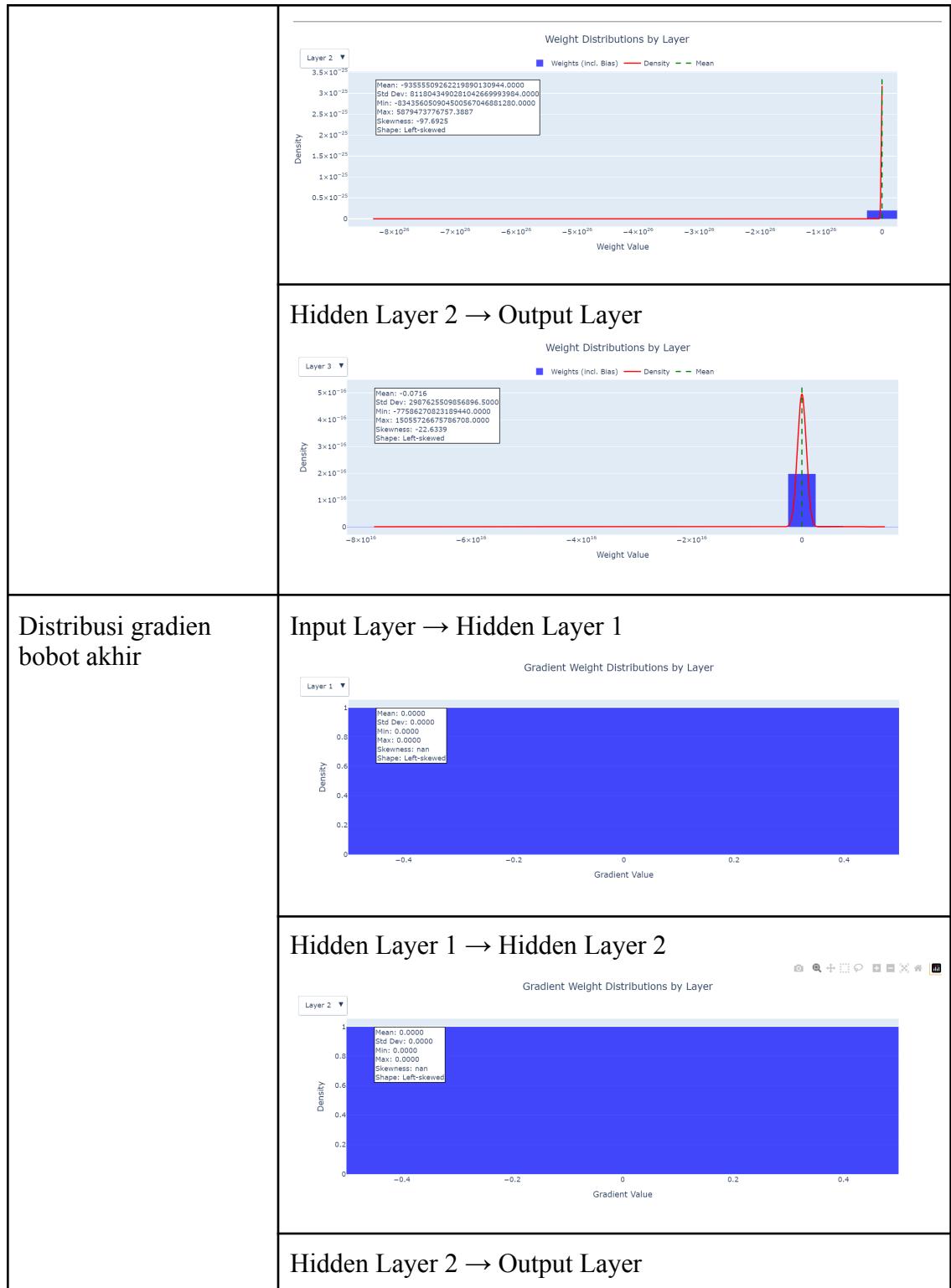


Distribusi bobot akhir

Input Layer → Hidden Layer 1

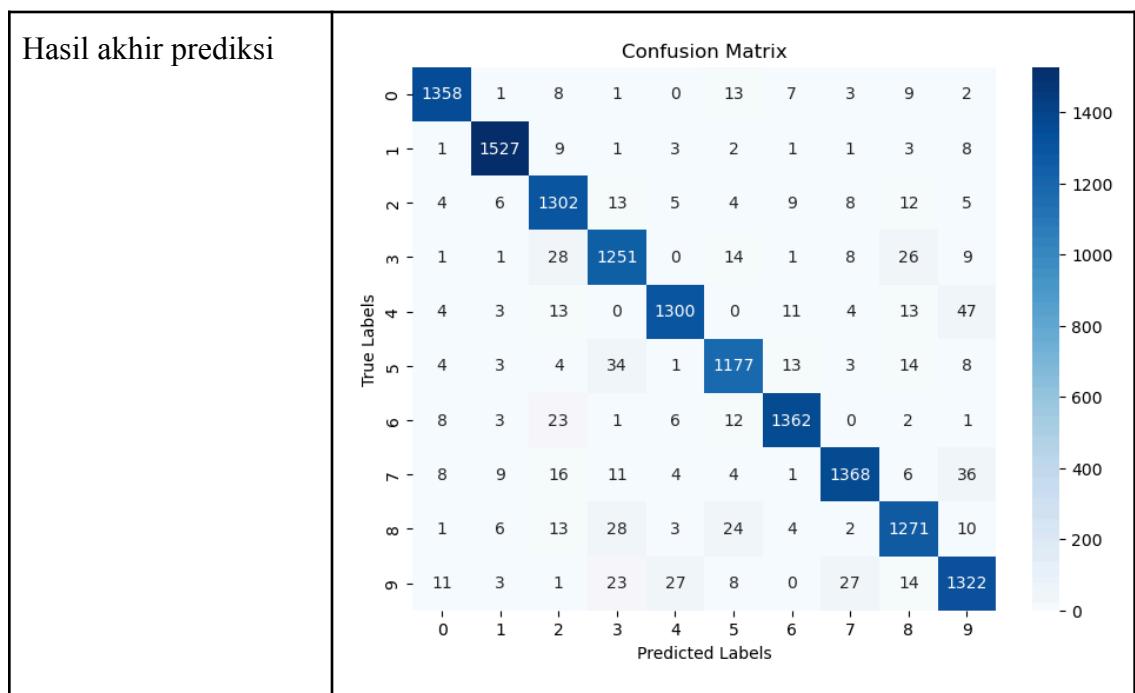


Hidden Layer 1 → Hidden Layer 2





2.2.4.5. Uniform Xavier Initialization



Skor validasi

Weighted F1-Score: 0.9456
Macro F1-Score: 0.9450
Cohen's Kappa: 0.9395
ROC-AUC (Weighted): 0.9965

Validation Log Loss: 0.2436

Classification Report:

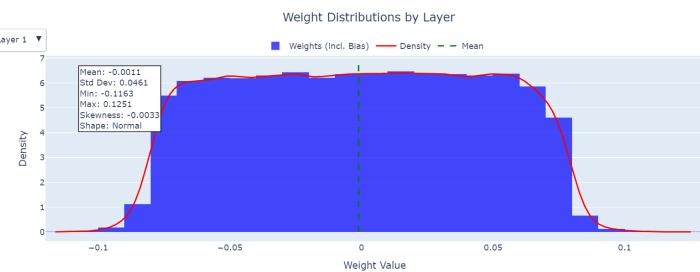
	precision	recall	f1-score	support
0	0.97	0.97	0.97	1402
1	0.98	0.98	0.98	1556
2	0.92	0.95	0.94	1368
3	0.92	0.93	0.93	1339
4	0.96	0.93	0.95	1395
5	0.94	0.93	0.93	1261
6	0.97	0.96	0.96	1418
7	0.96	0.94	0.95	1463
8	0.93	0.93	0.93	1362
9	0.91	0.92	0.92	1436
accuracy			0.95	14000
macro avg	0.95	0.95	0.95	14000
weighted avg	0.95	0.95	0.95	14000

Grafik loss

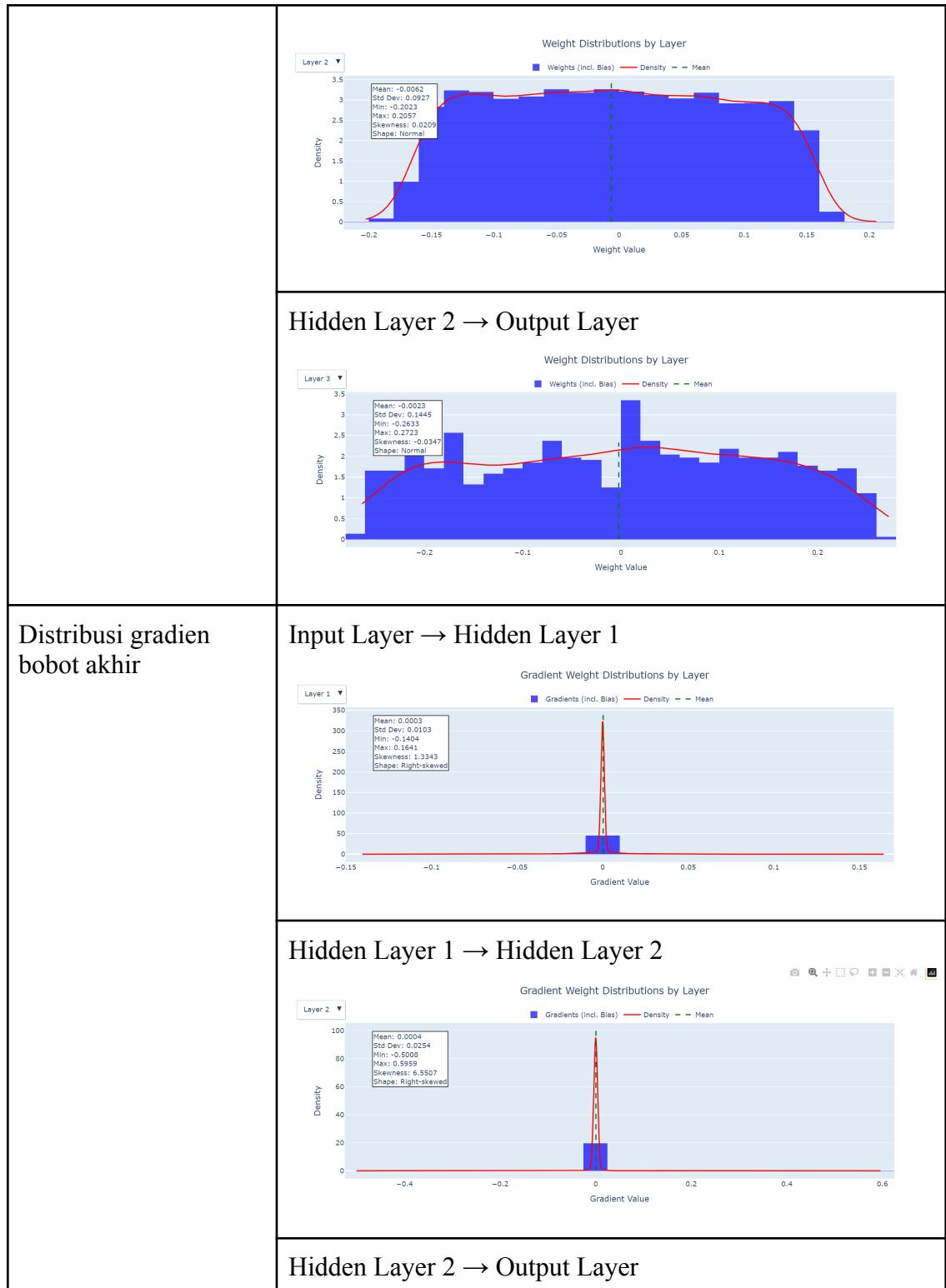


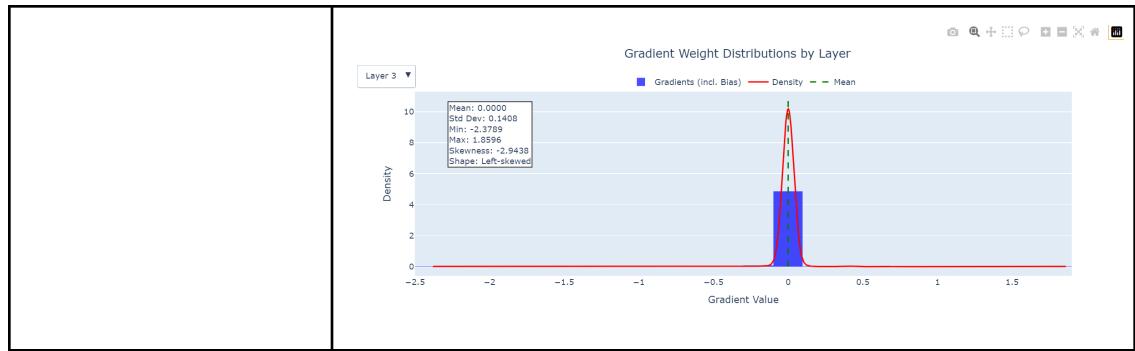
Distribusi bobot akhir

Input Layer → Hidden Layer 1

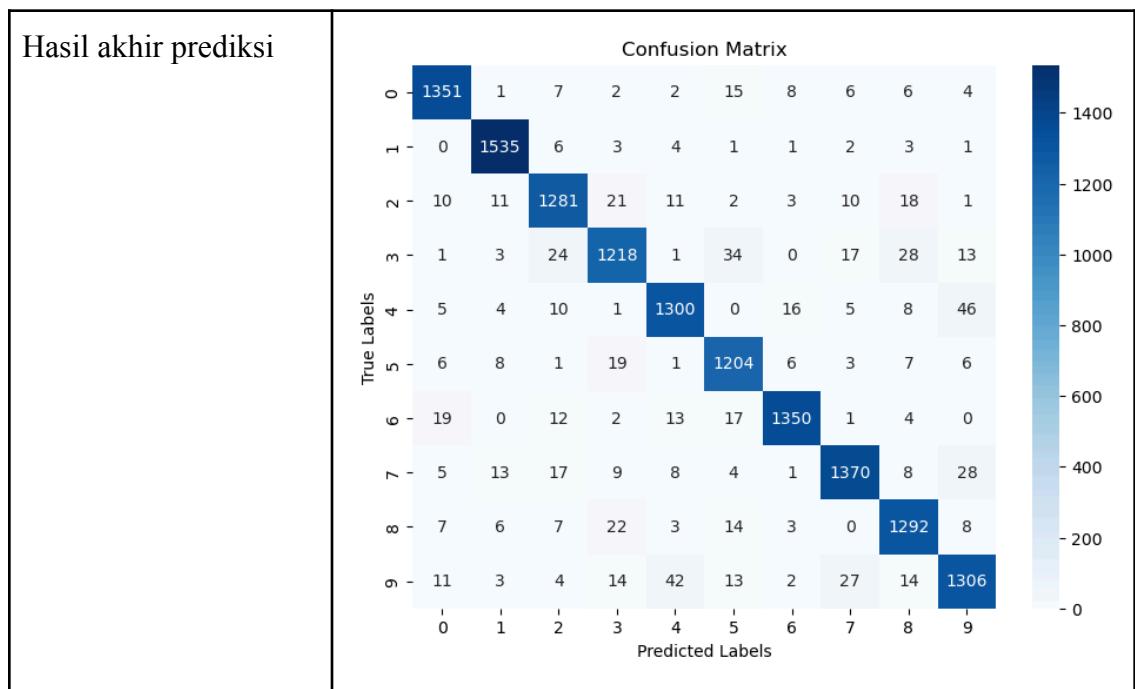


Hidden Layer 1 → Hidden Layer 2





2.2.4.6. Normal He Initialization



Skor validasi

```
Weighted F1-Score: 0.9433  
Macro F1-Score: 0.9428  
Cohen's Kappa: 0.9370  
ROC-AUC (Weighted): 0.9961
```

```
Validation Log Loss: 0.2458
```

Classification Report:

	precision	recall	f1-score	support
0	0.95	0.96	0.96	1402
1	0.97	0.99	0.98	1556
2	0.94	0.94	0.94	1368
3	0.93	0.91	0.92	1339
4	0.94	0.93	0.94	1395
5	0.92	0.95	0.94	1261
6	0.97	0.95	0.96	1418
7	0.95	0.94	0.94	1463
8	0.93	0.95	0.94	1362
9	0.92	0.91	0.92	1436
accuracy			0.94	14000
macro avg	0.94	0.94	0.94	14000
weighted avg	0.94	0.94	0.94	14000

Grafik loss

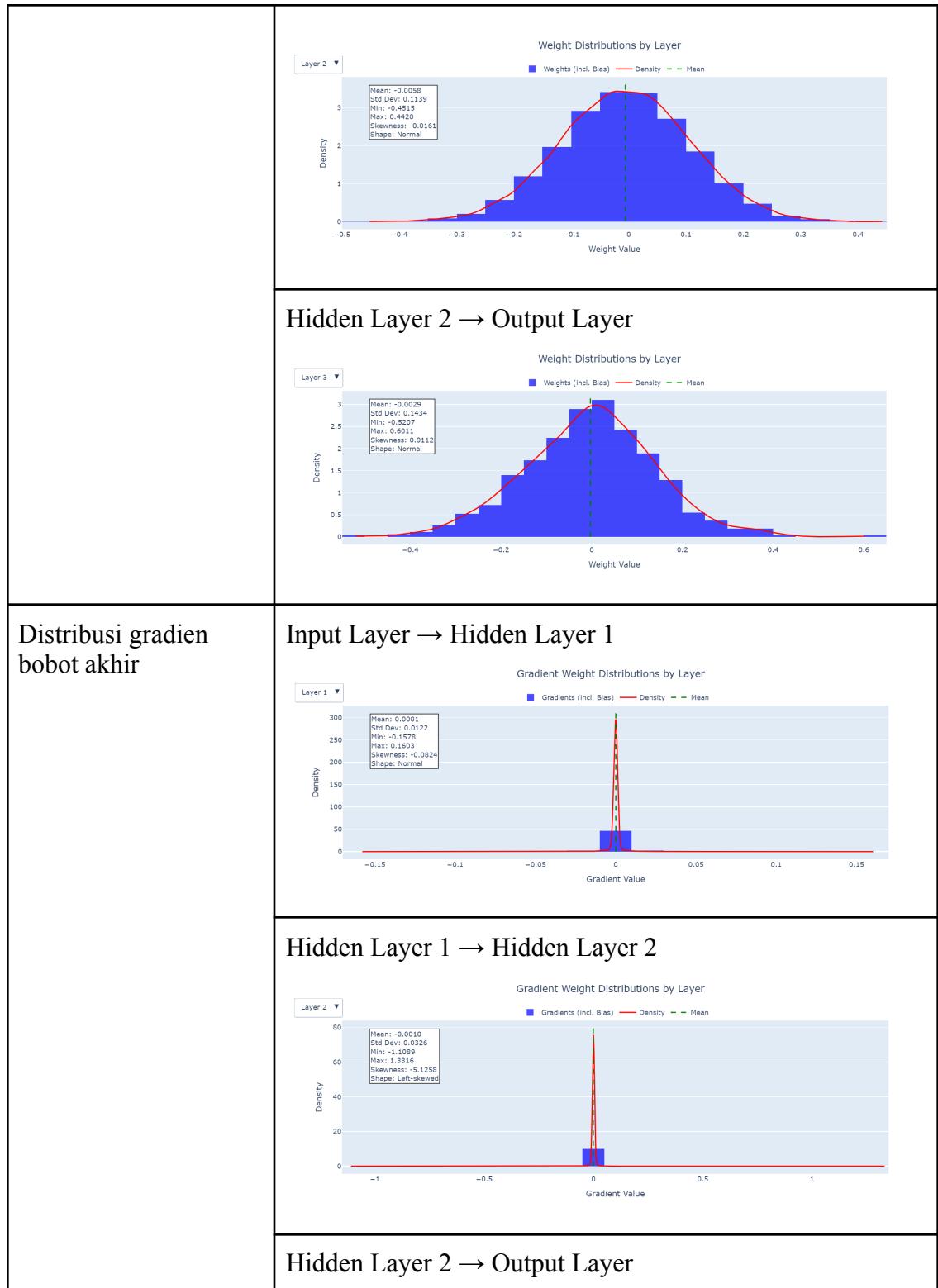


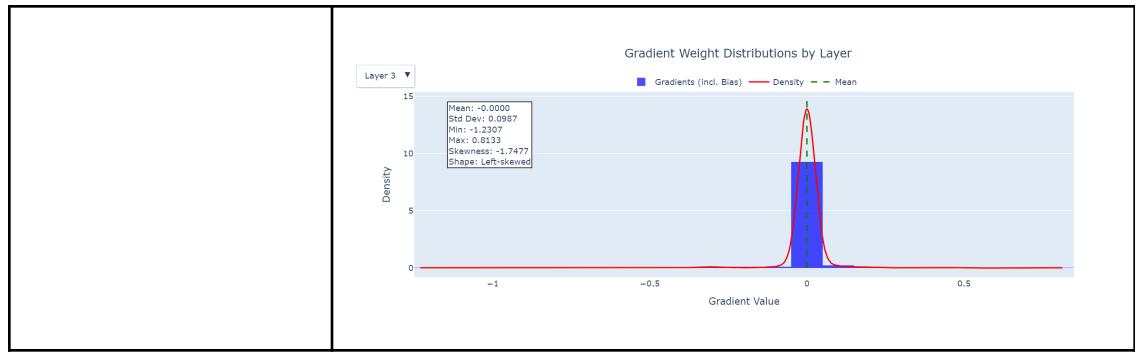
Distribusi bobot akhir

Input Layer → Hidden Layer 1

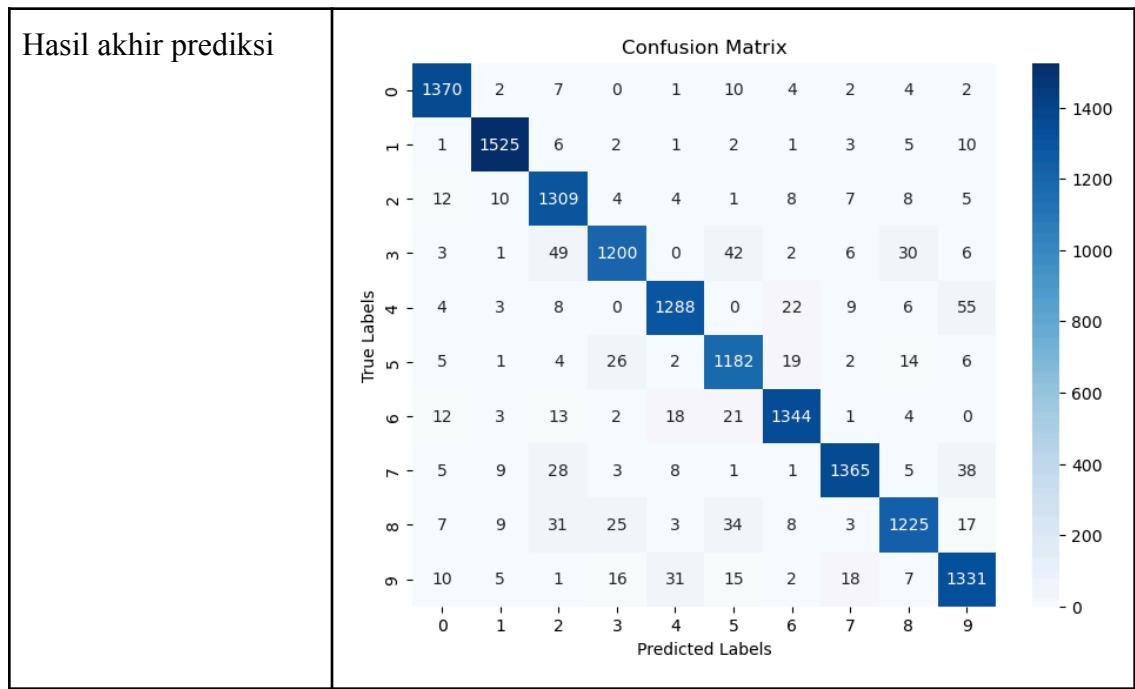


Hidden Layer 1 → Hidden Layer 2





2.2.4.7. Uniform He Initialization



Skor validasi

Weighted F1-Score: 0.9385
Macro F1-Score: 0.9377
Cohen's Kappa: 0.9316
ROC-AUC (Weighted): 0.9960

Validation Log Loss: 0.2513

Classification Report:

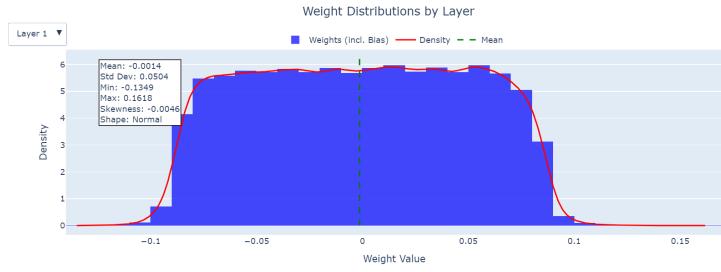
	precision	recall	f1-score	support
0	0.96	0.98	0.97	1402
1	0.97	0.98	0.98	1556
2	0.90	0.96	0.93	1368
3	0.94	0.90	0.92	1339
4	0.95	0.92	0.94	1395
5	0.90	0.94	0.92	1261
6	0.95	0.95	0.95	1418
7	0.96	0.93	0.95	1463
8	0.94	0.90	0.92	1362
9	0.91	0.93	0.92	1436
accuracy			0.94	14000
macro avg	0.94	0.94	0.94	14000
weighted avg	0.94	0.94	0.94	14000

Grafik loss

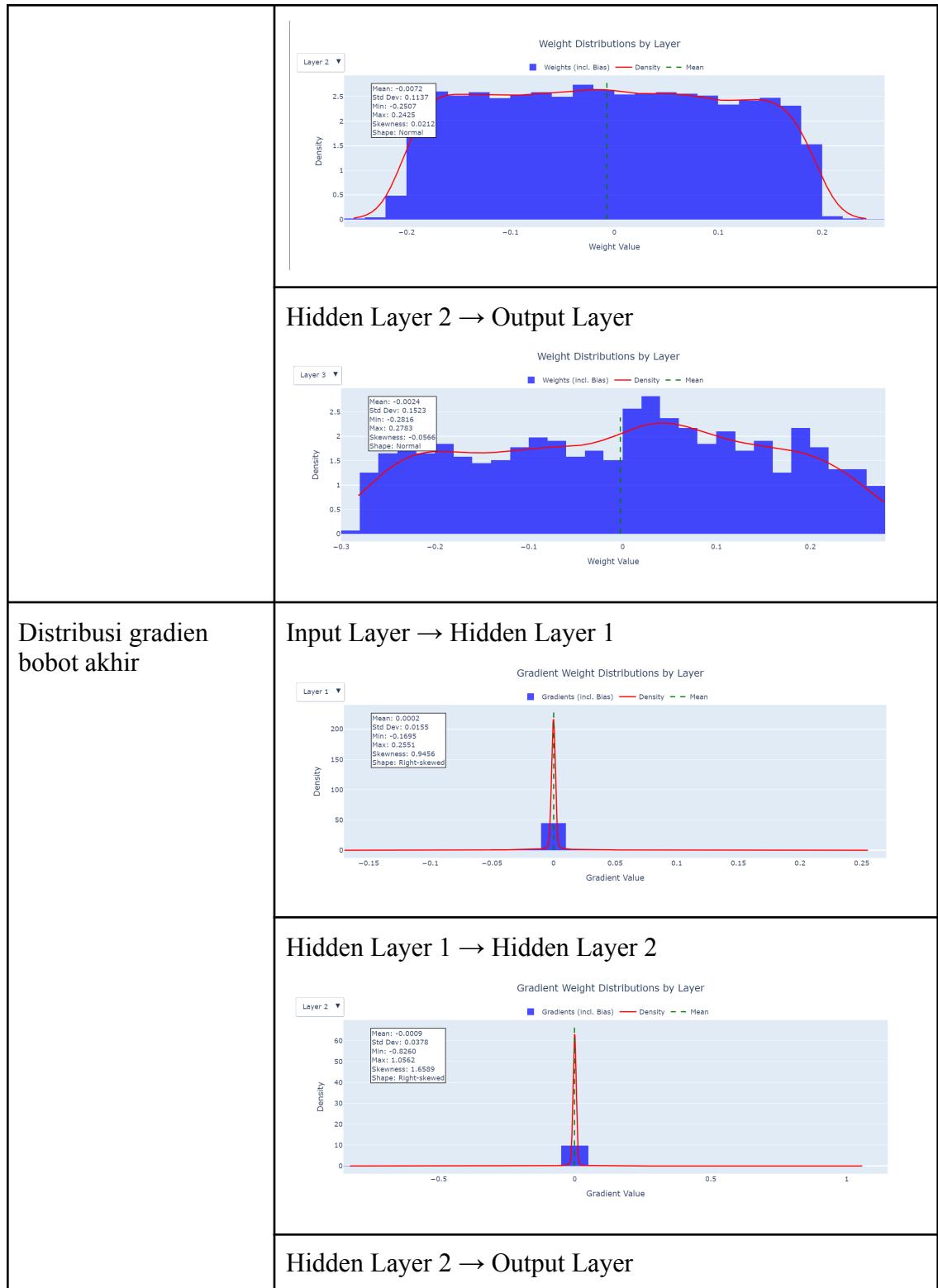


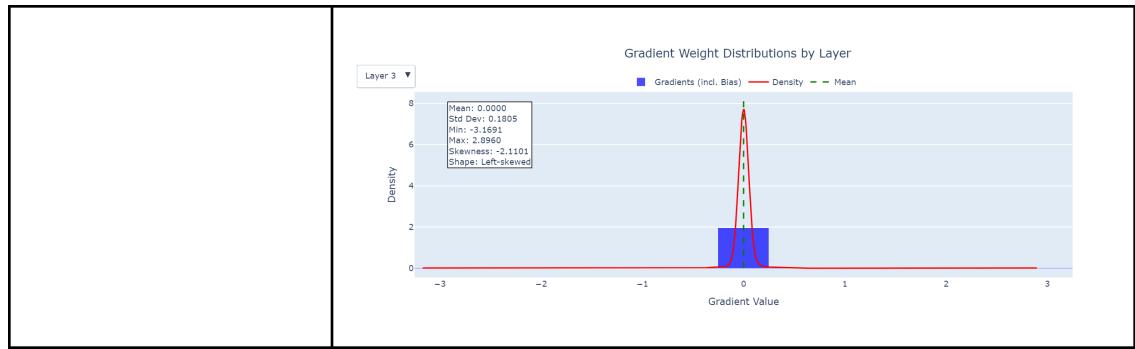
Distribusi bobot akhir

Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2





Dari hasil pengujian variasi bobot, didapatkan bahwa penggunaan Normal He *initialization method* memberikan hasil akurasi terbaik dibandingkan metode inisialisasi bobot lainnya. Faktor yang berpengaruh juga terhadap hasil yang didapat ini adalah penggunaan fungsi aktivasi ReLU oleh model yang digunakan untuk melakukan pengujian. Normal He *initialization method* merupakan metode terbaik untuk digunakan bersamaan dengan fungsi aktivasi ReLU karena metode ini menjaga besaran variansi dari bobot yang membuat model terhindar dari nonaktifnya neuron pada proses training yang sedang dilakukan akibat terlalu kecilnya bobot dari neuron.

2.2.6. Pengaruh Regularisasi

Parameter kontrol yang digunakan untuk membandingkan pengaruh antara penggunaan regularisasi adalah :

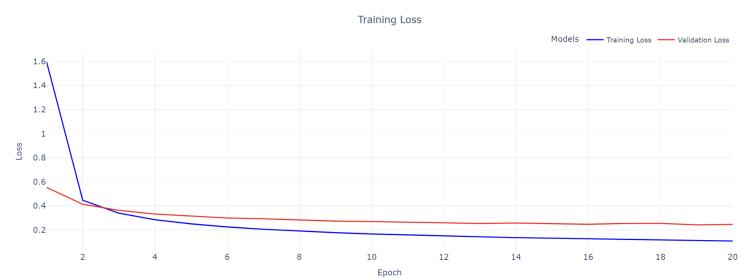
Jumlah layer	: 4 (input layer, hidden layer 1, hidden layer 2, output layer)
Neuron tiap layer	: 784 neuron, 150 neuron, 75 neuron, 10 neuron
Fungsi aktivasi (output layer)	: ReLU (hidden layer 1), ReLU (hidden layer 2), softmax
Inisialisasi Bobot	: Normal He Initialization
Fungsi loss	: Categorical Cross-Entropy
Batch size	: 64
Learning rate	: 0.001
Epoch maksimal	: 20

2.2.6.1. Tanpa Regularisasi

Hasil akhir prediksi		Confusion Matrix											
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9		
		1351	1	7	2	2	15	8	6	6	4	- 1400	- 1200
		0	1535	6	3	4	1	1	2	3	1	- 1000	- 800
		10	11	1281	21	11	2	3	10	18	1	- 600	- 400
		1	3	24	1218	1	34	0	17	28	13	- 200	0
		5	4	10	1	1300	0	16	5	8	46		
		6	8	1	19	1	1204	6	3	7	6		
		19	0	12	2	13	17	1350	1	4	0		
		5	13	17	9	8	4	1	1370	8	28		
		8	6	7	22	3	14	3	0	1292	8		
		9	3	4	14	42	13	2	27	14	1306		
		0	1	2	3	4	5	6	7	8	9		

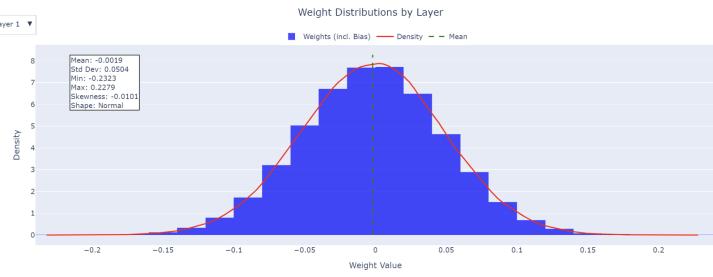
Skor validasi		Weighted F1-Score: 0.9433				
		Macro F1-Score: 0.9428				
		Cohen's Kappa: 0.9370				
		ROC-AUC (Weighted): 0.9961				
		Validation Log Loss: 0.2461				
Classification Report:						
		precision	recall	f1-score	support	
	0	0.95	0.96	0.96	1402	
	1	0.97	0.99	0.98	1556	
	2	0.94	0.94	0.94	1368	
	3	0.93	0.91	0.92	1339	
	4	0.94	0.93	0.94	1395	
	5	0.92	0.95	0.94	1261	
	6	0.97	0.95	0.96	1418	
	7	0.95	0.94	0.94	1463	
	8	0.93	0.95	0.94	1362	
	9	0.92	0.91	0.92	1436	
	accuracy				0.94	14000
	macro avg				0.94	14000
	weighted avg				0.94	14000

Grafik loss

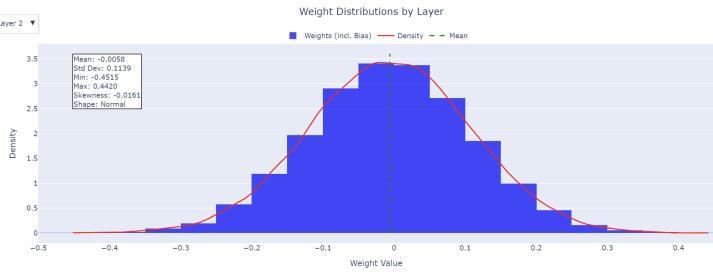


Distribusi bobot akhir

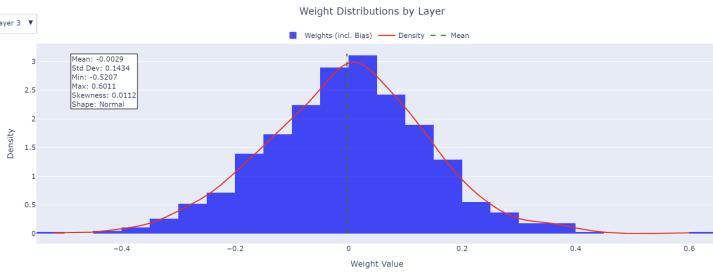
Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2

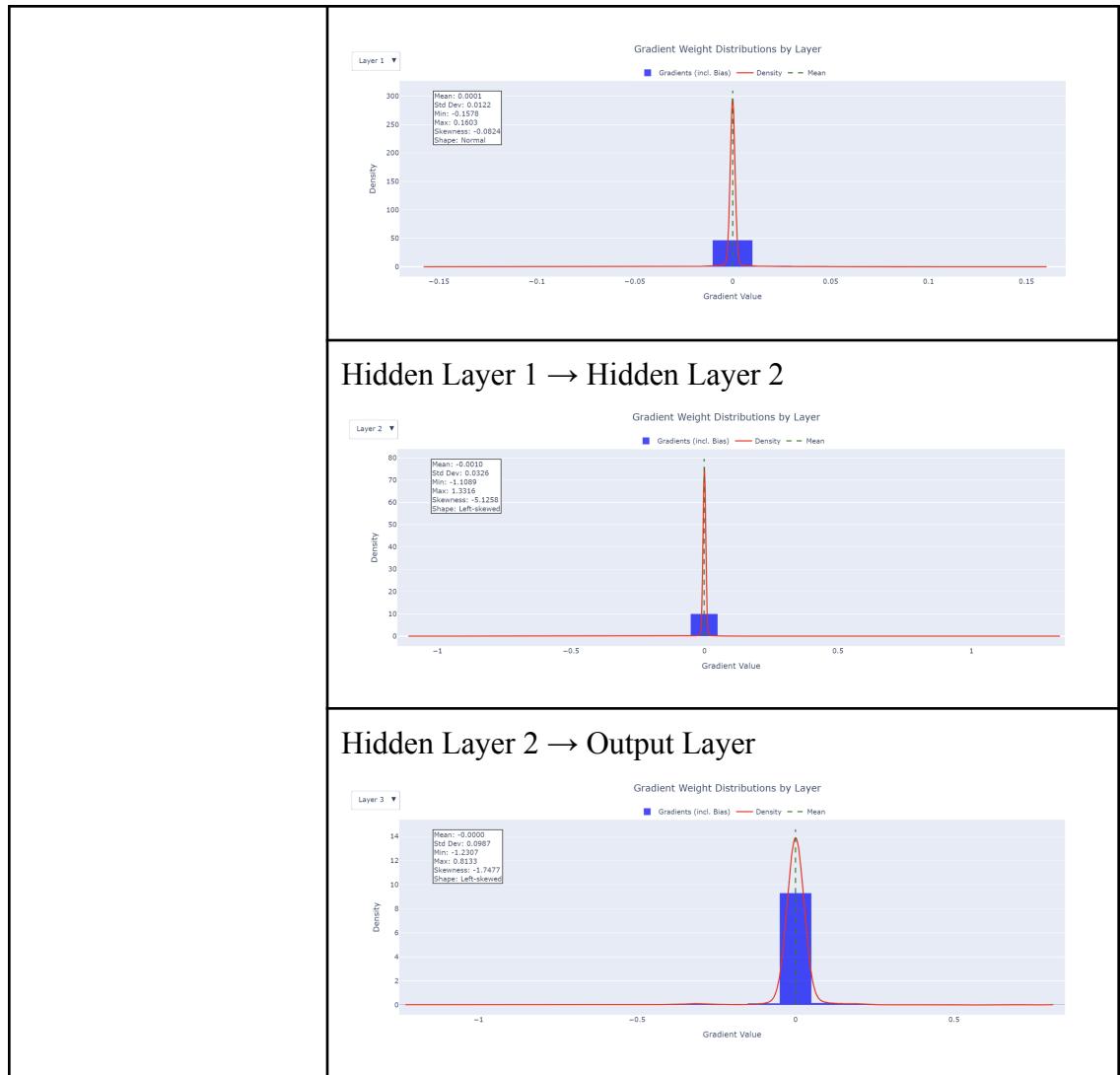


Hidden Layer 2 → Output Layer



Distribusi gradien bobot akhir

Input Layer → Hidden Layer 1



2.2.6.2. Dengan Regularisasi L1

Hasil akhir prediksi		Confusion Matrix											
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9	0	
		1365	0	8	1	0	11	8	1	8	0	1400	
		1	1517	14	8	0	0	1	1	13	1	1200	
		2	6	5	1305	3	7	1	17	12	10	2	1000
		3	9	1	39	1219	1	22	7	15	15	11	800
		4	3	2	6	1	1320	2	12	5	10	34	600
		5	5	5	7	12	1	1200	21	1	3	6	400
		6	12	2	9	1	10	16	1364	0	4	0	200
		7	15	15	27	4	11	6	1	1356	2	26	0
		8	4	13	13	13	2	15	14	5	1273	10	0
		9	12	5	1	23	51	19	1	29	14	1281	0
		0	1	2	3	4	5	6	7	8	9	9	0

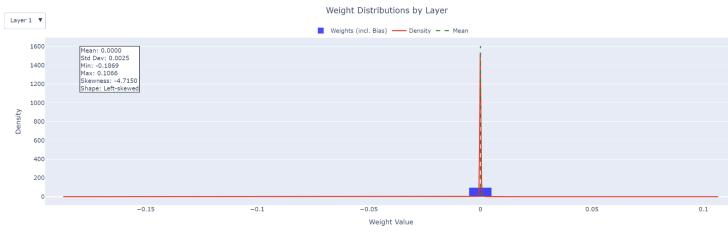
Skor validasi		Weighted F1-Score: 0.9427				
		Macro F1-Score: 0.9424				
		Cohen's Kappa: 0.9365				
		ROC-AUC (Weighted): 0.9968				
		Validation Log Loss: 0.2160				
Classification Report:						
		precision	recall	f1-score	support	
	0	0.95	0.97	0.96	1402	
	1	0.97	0.97	0.97	1556	
	2	0.91	0.95	0.93	1368	
	3	0.95	0.91	0.93	1339	
	4	0.94	0.95	0.94	1395	
	5	0.93	0.95	0.94	1261	
	6	0.94	0.96	0.95	1418	
	7	0.95	0.93	0.94	1463	
	8	0.94	0.93	0.94	1362	
	9	0.93	0.89	0.91	1436	
	accuracy			0.94	14000	
	macro avg		0.94	0.94	0.94	14000
	weighted avg		0.94	0.94	0.94	14000

Grafik loss

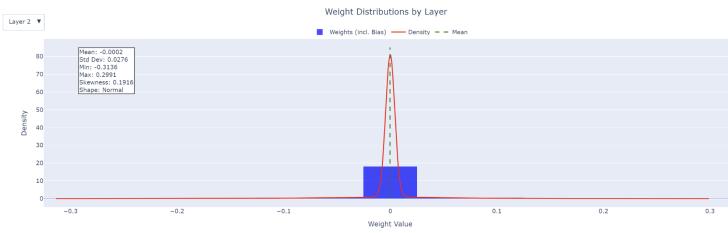


Distribusi bobot akhir

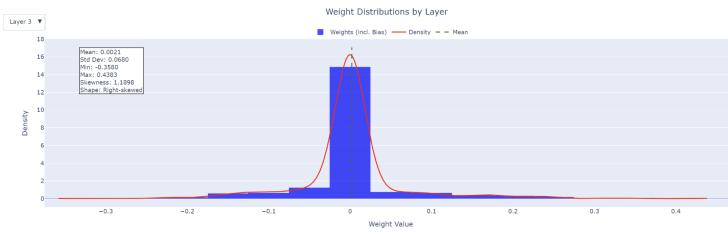
Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2

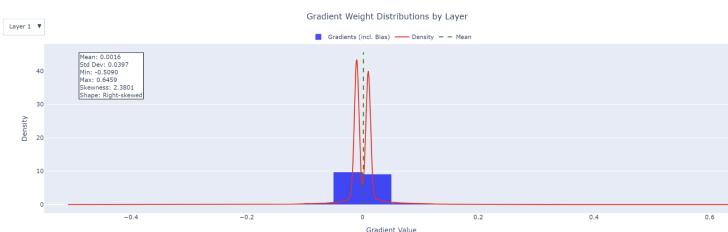


Hidden Layer 2 → Output Layer

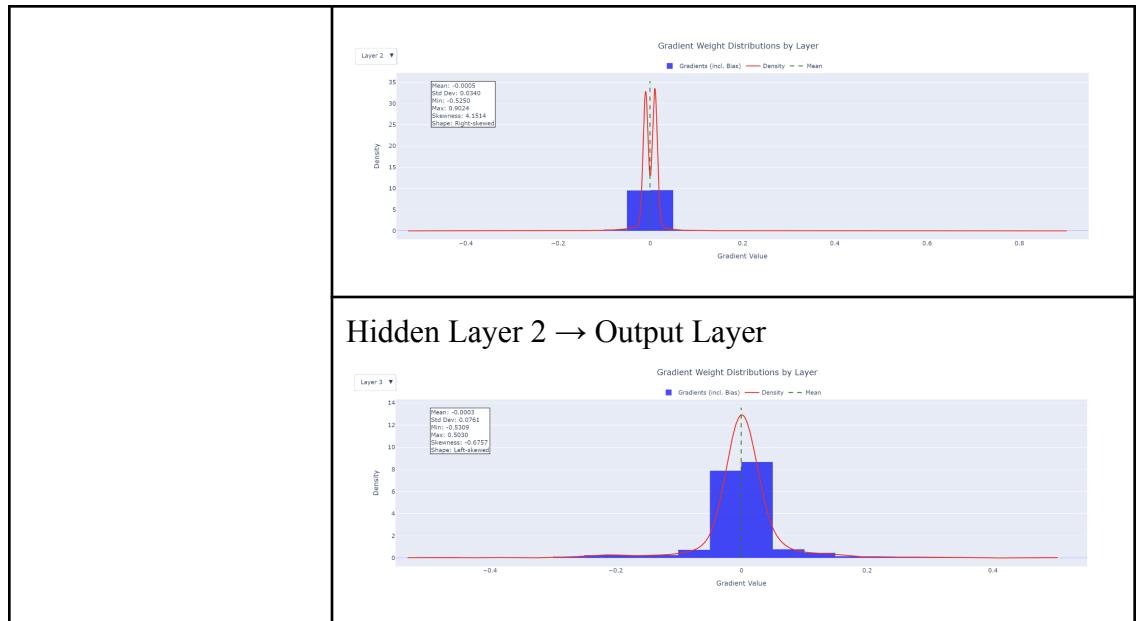


Distribusi gradien bobot akhir

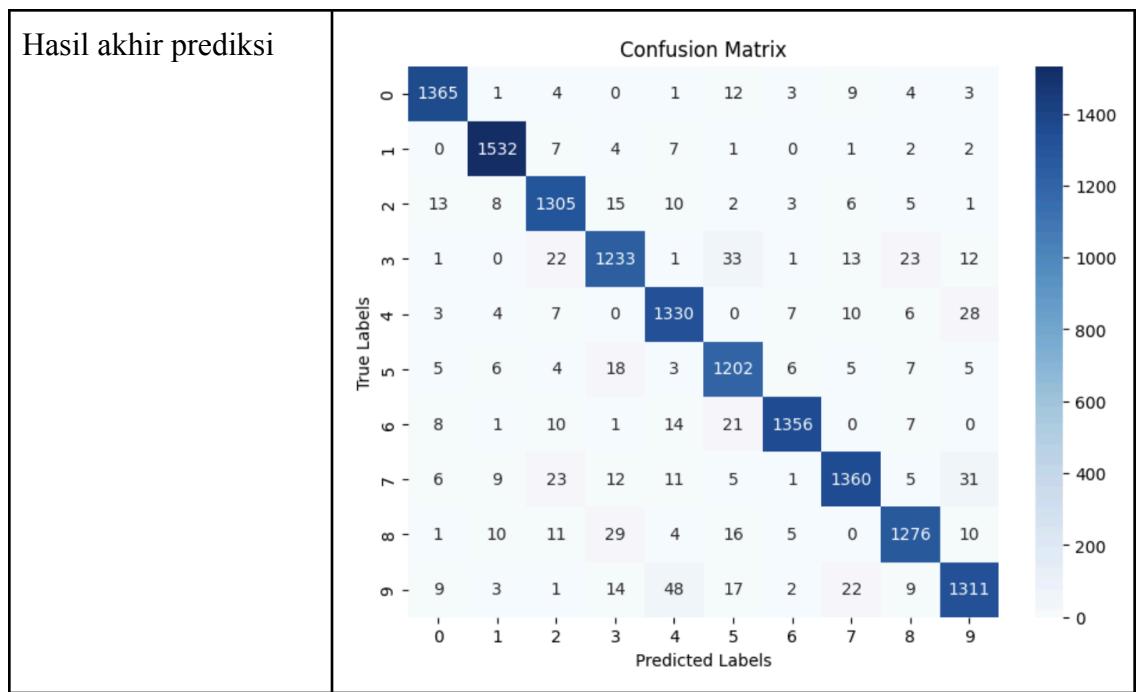
Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2



2.2.6.3. Dengan Regularisasi L2



Skor validasi

Weighted F1-Score: 0.9478

Macro F1-Score: 0.9473

Cohen's Kappa: 0.9420

ROC-AUC (Weighted): 0.9972

Validation Log Loss: 0.2016

Classification Report:

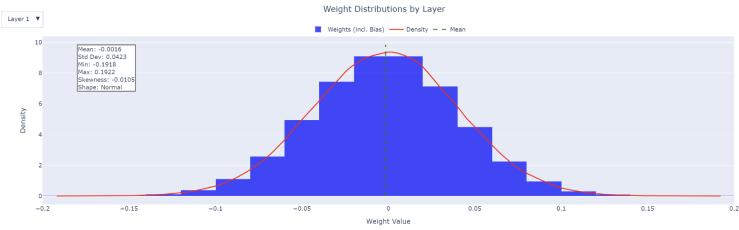
	precision	recall	f1-score	support
0	0.97	0.97	0.97	1402
1	0.97	0.98	0.98	1556
2	0.94	0.95	0.94	1368
3	0.93	0.92	0.93	1339
4	0.93	0.95	0.94	1395
5	0.92	0.95	0.94	1261
6	0.98	0.96	0.97	1418
7	0.95	0.93	0.94	1463
8	0.95	0.94	0.94	1362
9	0.93	0.91	0.92	1436
accuracy			0.95	14000
macro avg	0.95	0.95	0.95	14000
weighted avg	0.95	0.95	0.95	14000

Grafik loss

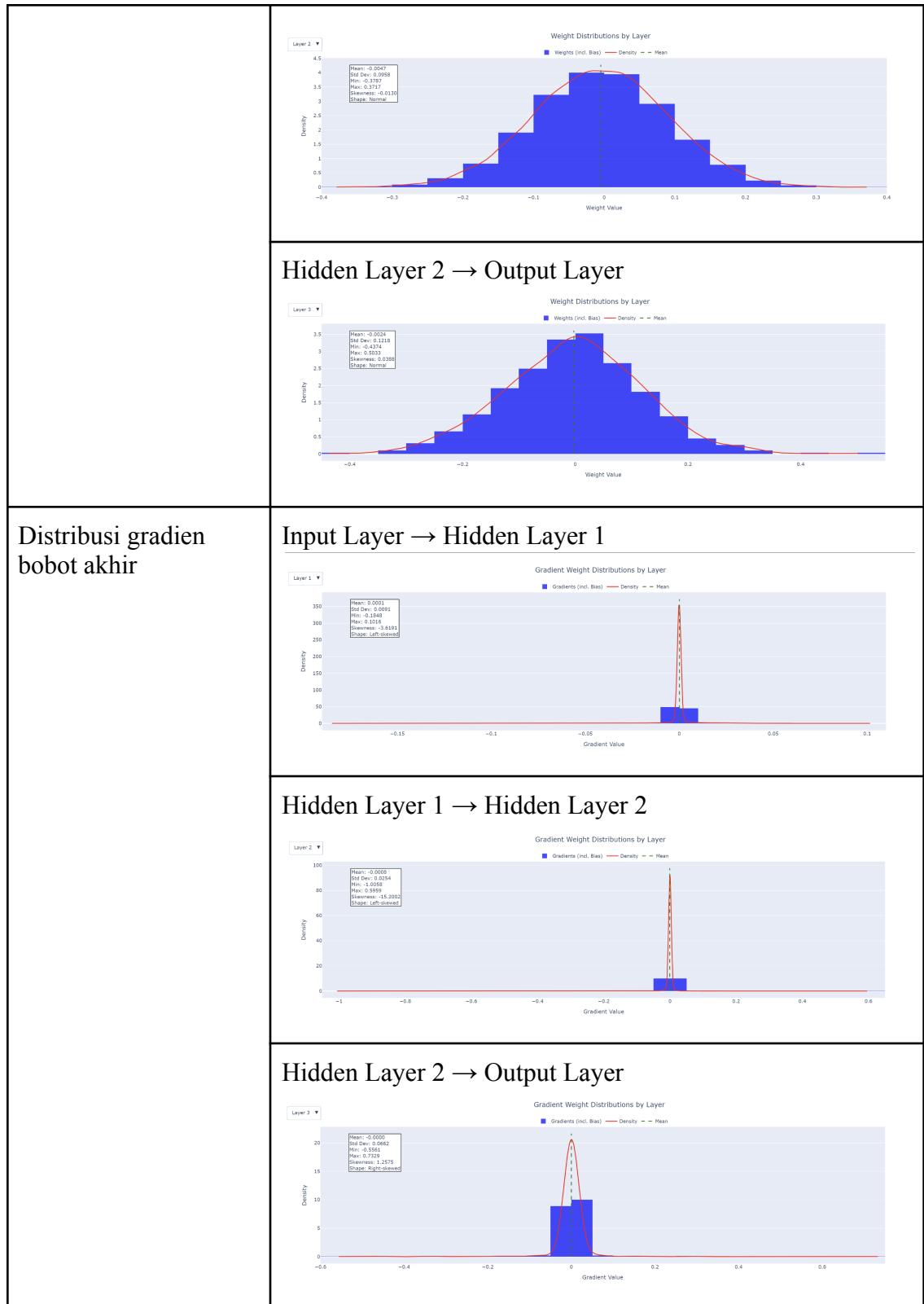


Distribusi bobot akhir

Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2



Berdasarkan hasil dari perbandingan hasil prediksi antara model tanpa regularisasi, model dengan regularisasi L1, serta model dengan regularisasi L2,

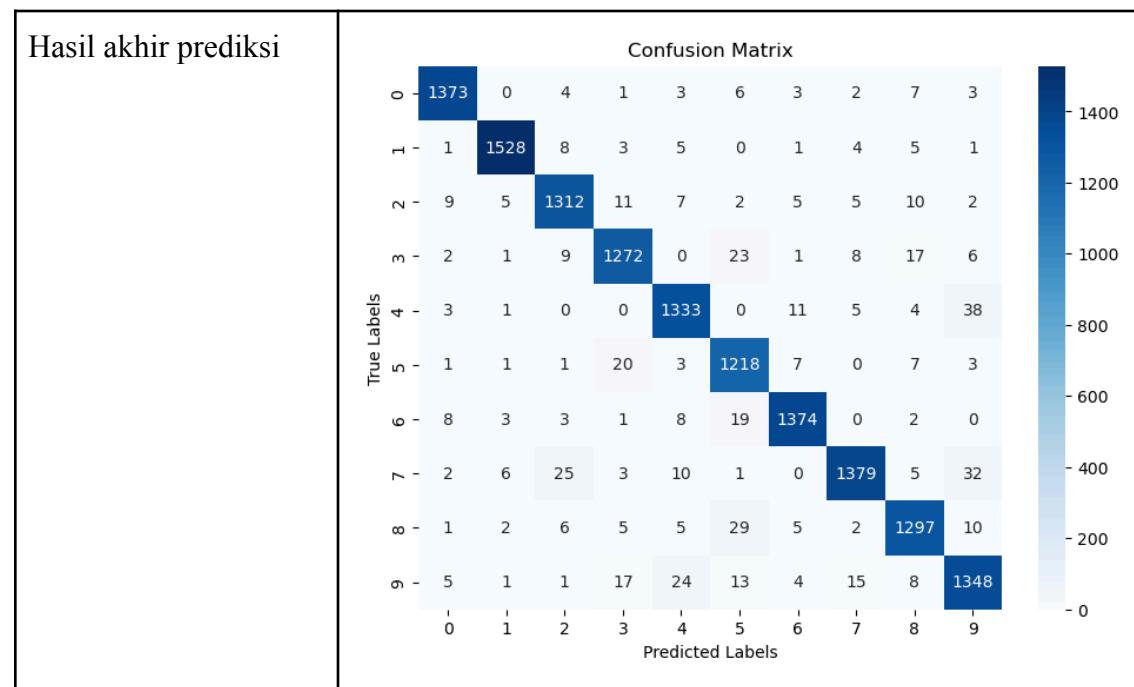
didapatkan bahwa model dengan regularisasi L2 memiliki hasil prediksi yang terbaik.

2.2.7. Perbandingan dengan Library Sklearn

Parameter yang digunakan untuk test menggunakan library Sklearn adalah :

Jumlah layer	: 4 (input layer, hidden layer 1, hidden layer 2, output layer)
Neuron tiap layer	: 784 neuron, 150 neuron, 75 neuron, 10 neuron
Fungsi aktivasi	: ReLU (hidden layer 1), ReLU (hidden layer 2), softmax (output layer)
Inisialisasi Bobot	: Uniform Xavier Initialization
Fungsi loss	: Categorical Cross-Entropy
Batch size	: 64
Epoch maksimal	: 20
Learning rate	: 0.001

2.2.6.4. Library Sklearn MLP



Skor validasi

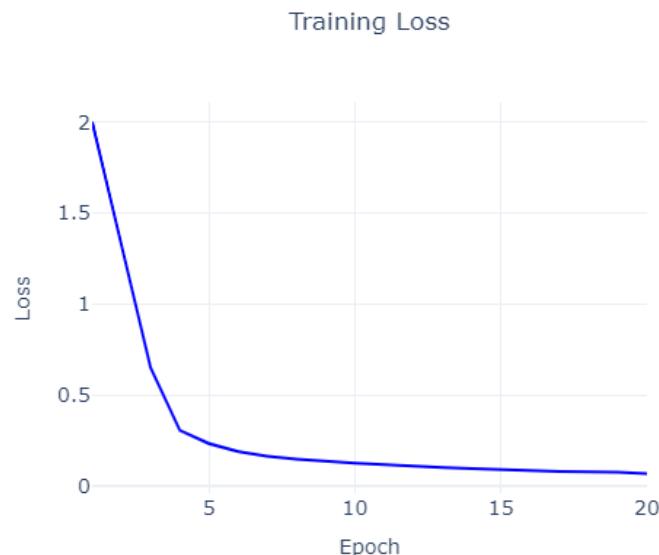
```
Weighted F1-Score: 0.9596
Macro F1-Score: 0.9592
Cohen's Kappa: 0.9551
ROC-AUC (Weighted): 0.9976

Validation Log Loss: 0.1754

Classification Report:
precision    recall    f1-score   support
          0       0.98      0.98      0.98     1402
          1       0.99      0.98      0.98     1556
          2       0.96      0.96      0.96     1368
          3       0.95      0.95      0.95     1339
          4       0.95      0.96      0.95     1395
          5       0.93      0.97      0.95     1261
          6       0.97      0.97      0.97     1418
          7       0.97      0.94      0.96     1463
          8       0.95      0.95      0.95     1362
          9       0.93      0.94      0.94     1436

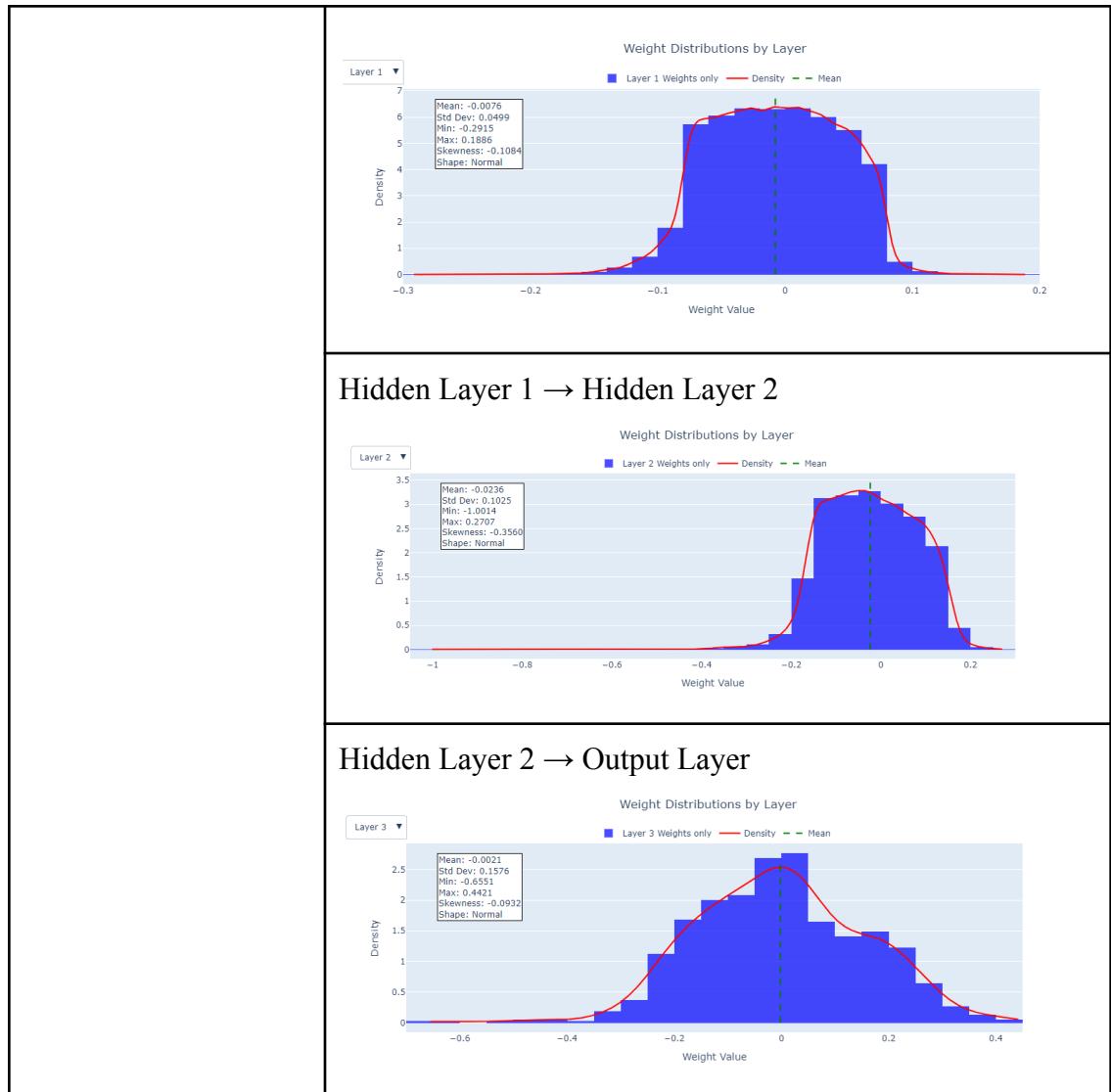
accuracy                           0.96      14000
macro avg                           0.96      0.96      0.96     14000
weighted avg                        0.96      0.96      0.96     14000
```

Grafik loss



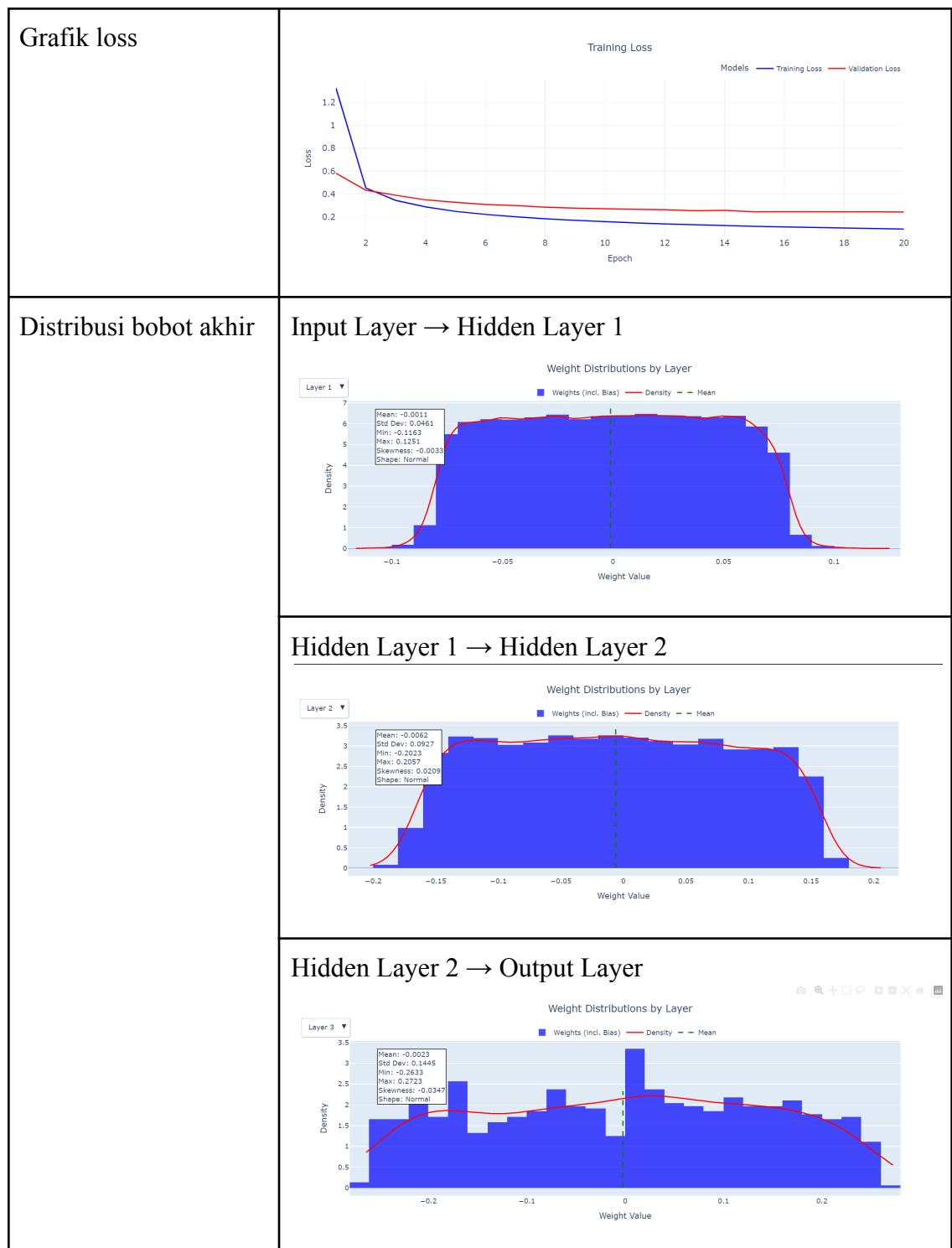
Distribusi bobot akhir

Input Layer → Hidden Layer 1



2.2.6.5. FFNN Buatan

Hasil akhir prediksi	<p style="text-align: center;">Confusion Matrix</p> <table border="1" style="margin-top: 10px; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2"></th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> </tr> <tr> <th rowspan="2">True Labels</th> <th>Predicted Labels</th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1358</td> <td>1</td> <td>8</td> <td>1</td> <td>0</td> <td>13</td> <td>7</td> <td>3</td> <td>9</td> <td>2</td> </tr> <tr> <td>1</td> <td>1</td> <td>1527</td> <td>9</td> <td>1</td> <td>3</td> <td>2</td> <td>1</td> <td>1</td> <td>3</td> <td>8</td> </tr> <tr> <td>2</td> <td>4</td> <td>6</td> <td>1302</td> <td>13</td> <td>5</td> <td>4</td> <td>9</td> <td>8</td> <td>12</td> <td>5</td> </tr> <tr> <td>3</td> <td>1</td> <td>1</td> <td>28</td> <td>1251</td> <td>0</td> <td>14</td> <td>1</td> <td>8</td> <td>26</td> <td>9</td> </tr> <tr> <td>4</td> <td>4</td> <td>3</td> <td>13</td> <td>0</td> <td>1300</td> <td>0</td> <td>11</td> <td>4</td> <td>13</td> <td>47</td> </tr> <tr> <td>5</td> <td>4</td> <td>3</td> <td>4</td> <td>34</td> <td>1</td> <td>1177</td> <td>13</td> <td>3</td> <td>14</td> <td>8</td> </tr> <tr> <td>6</td> <td>8</td> <td>3</td> <td>23</td> <td>1</td> <td>6</td> <td>12</td> <td>1362</td> <td>0</td> <td>2</td> <td>1</td> </tr> <tr> <td>7</td> <td>8</td> <td>9</td> <td>16</td> <td>11</td> <td>4</td> <td>4</td> <td>1</td> <td>1368</td> <td>6</td> <td>36</td> </tr> <tr> <td>8</td> <td>1</td> <td>6</td> <td>13</td> <td>28</td> <td>3</td> <td>24</td> <td>4</td> <td>2</td> <td>1271</td> <td>10</td> </tr> <tr> <td>9</td> <td>11</td> <td>3</td> <td>1</td> <td>23</td> <td>27</td> <td>8</td> <td>0</td> <td>27</td> <td>14</td> <td>1322</td> </tr> <tr> <td></td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> </tbody> </table>			0	1	2	3	4	5	6	7	8	9	True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9	0	1358	1	8	1	0	13	7	3	9	2	1	1	1527	9	1	3	2	1	1	3	8	2	4	6	1302	13	5	4	9	8	12	5	3	1	1	28	1251	0	14	1	8	26	9	4	4	3	13	0	1300	0	11	4	13	47	5	4	3	4	34	1	1177	13	3	14	8	6	8	3	23	1	6	12	1362	0	2	1	7	8	9	16	11	4	4	1	1368	6	36	8	1	6	13	28	3	24	4	2	1271	10	9	11	3	1	23	27	8	0	27	14	1322		0	1	2	3	4	5	6	7	8	9
		0	1	2	3	4	5	6	7	8	9																																																																																																																																							
True Labels	Predicted Labels	0	1	2	3	4	5	6	7	8	9																																																																																																																																							
	0	1358	1	8	1	0	13	7	3	9	2																																																																																																																																							
1	1	1527	9	1	3	2	1	1	3	8																																																																																																																																								
2	4	6	1302	13	5	4	9	8	12	5																																																																																																																																								
3	1	1	28	1251	0	14	1	8	26	9																																																																																																																																								
4	4	3	13	0	1300	0	11	4	13	47																																																																																																																																								
5	4	3	4	34	1	1177	13	3	14	8																																																																																																																																								
6	8	3	23	1	6	12	1362	0	2	1																																																																																																																																								
7	8	9	16	11	4	4	1	1368	6	36																																																																																																																																								
8	1	6	13	28	3	24	4	2	1271	10																																																																																																																																								
9	11	3	1	23	27	8	0	27	14	1322																																																																																																																																								
	0	1	2	3	4	5	6	7	8	9																																																																																																																																								
Skor validasi	<p>Weighted F1-Score: 0.9456 Macro F1-Score: 0.9450 Cohen's Kappa: 0.9395 ROC-AUC (Weighted): 0.9965</p> <p>Validation Log Loss: 0.2436</p> <p>Classification Report:</p> <table border="1" style="margin-top: 10px; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.97</td> <td>0.97</td> <td>0.97</td> <td>1402</td> </tr> <tr> <td>1</td> <td>0.98</td> <td>0.98</td> <td>0.98</td> <td>1556</td> </tr> <tr> <td>2</td> <td>0.92</td> <td>0.95</td> <td>0.94</td> <td>1368</td> </tr> <tr> <td>3</td> <td>0.92</td> <td>0.93</td> <td>0.93</td> <td>1339</td> </tr> <tr> <td>4</td> <td>0.96</td> <td>0.93</td> <td>0.95</td> <td>1395</td> </tr> <tr> <td>5</td> <td>0.94</td> <td>0.93</td> <td>0.93</td> <td>1261</td> </tr> <tr> <td>6</td> <td>0.97</td> <td>0.96</td> <td>0.96</td> <td>1418</td> </tr> <tr> <td>7</td> <td>0.96</td> <td>0.94</td> <td>0.95</td> <td>1463</td> </tr> <tr> <td>8</td> <td>0.93</td> <td>0.93</td> <td>0.93</td> <td>1362</td> </tr> <tr> <td>9</td> <td>0.91</td> <td>0.92</td> <td>0.92</td> <td>1436</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.95</td> <td>14000</td> </tr> <tr> <td>macro avg</td> <td>0.95</td> <td>0.95</td> <td>0.95</td> <td>14000</td> </tr> <tr> <td>weighted avg</td> <td>0.95</td> <td>0.95</td> <td>0.95</td> <td>14000</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.97	0.97	0.97	1402	1	0.98	0.98	0.98	1556	2	0.92	0.95	0.94	1368	3	0.92	0.93	0.93	1339	4	0.96	0.93	0.95	1395	5	0.94	0.93	0.93	1261	6	0.97	0.96	0.96	1418	7	0.96	0.94	0.95	1463	8	0.93	0.93	0.93	1362	9	0.91	0.92	0.92	1436	accuracy			0.95	14000	macro avg	0.95	0.95	0.95	14000	weighted avg	0.95	0.95	0.95	14000																																																																											
	precision	recall	f1-score	support																																																																																																																																														
0	0.97	0.97	0.97	1402																																																																																																																																														
1	0.98	0.98	0.98	1556																																																																																																																																														
2	0.92	0.95	0.94	1368																																																																																																																																														
3	0.92	0.93	0.93	1339																																																																																																																																														
4	0.96	0.93	0.95	1395																																																																																																																																														
5	0.94	0.93	0.93	1261																																																																																																																																														
6	0.97	0.96	0.96	1418																																																																																																																																														
7	0.96	0.94	0.95	1463																																																																																																																																														
8	0.93	0.93	0.93	1362																																																																																																																																														
9	0.91	0.92	0.92	1436																																																																																																																																														
accuracy			0.95	14000																																																																																																																																														
macro avg	0.95	0.95	0.95	14000																																																																																																																																														
weighted avg	0.95	0.95	0.95	14000																																																																																																																																														



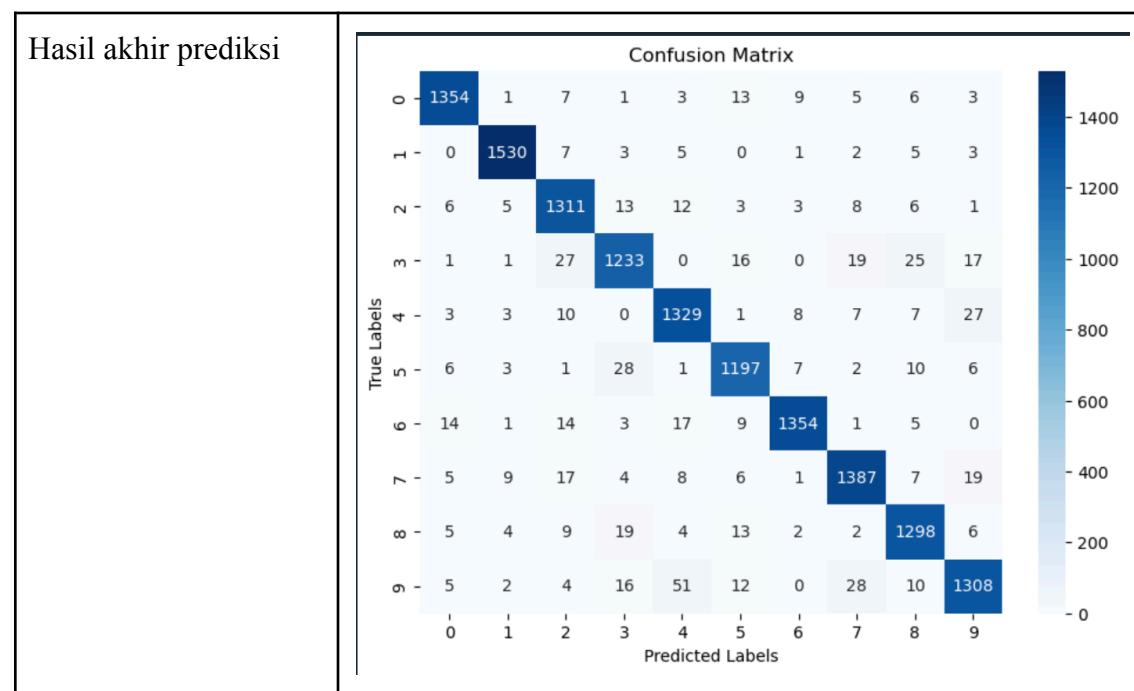
2.2.8. Pengujian terhadap Dataset MNIST_784

Parameter yang digunakan untuk test menggunakan mnist_784 adalah :

Jumlah layer : 4 (input layer, hidden layer 1, hidden layer 2, output layer)

Neuron tiap layer	: 784 neuron, 150 neuron, 75 neuron, 10 neuron
Fungsi aktivasi	: ReLU (hidden layer 1), ReLU (hidden layer 2), softmax
(output layer)	
Inisialisasi Bobot	: Normal He Initialization
Fungsi loss	: Categorical Cross-Entropy
Batch size	: 64
Epoch maksimal	: 50
Learning rate	: 0.001

2.2.6.1 dataset mnist_784



Skor validasi

Weighted F1-Score: 0.9501
Macro F1-Score: 0.9496
Cohen's Kappa: 0.9445
ROC-AUC (Weighted): 0.9968

Validation Log Loss: 0.2667

Classification Report:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	1402
1	0.98	0.98	0.98	1556
2	0.93	0.96	0.94	1368
3	0.93	0.92	0.93	1339
4	0.93	0.95	0.94	1395
5	0.94	0.95	0.95	1261
6	0.98	0.95	0.97	1418
7	0.95	0.95	0.95	1463
8	0.94	0.95	0.95	1362
9	0.94	0.91	0.93	1436
accuracy			0.95	14000
macro avg	0.95	0.95	0.95	14000
weighted avg	0.95	0.95	0.95	14000

Grafik loss

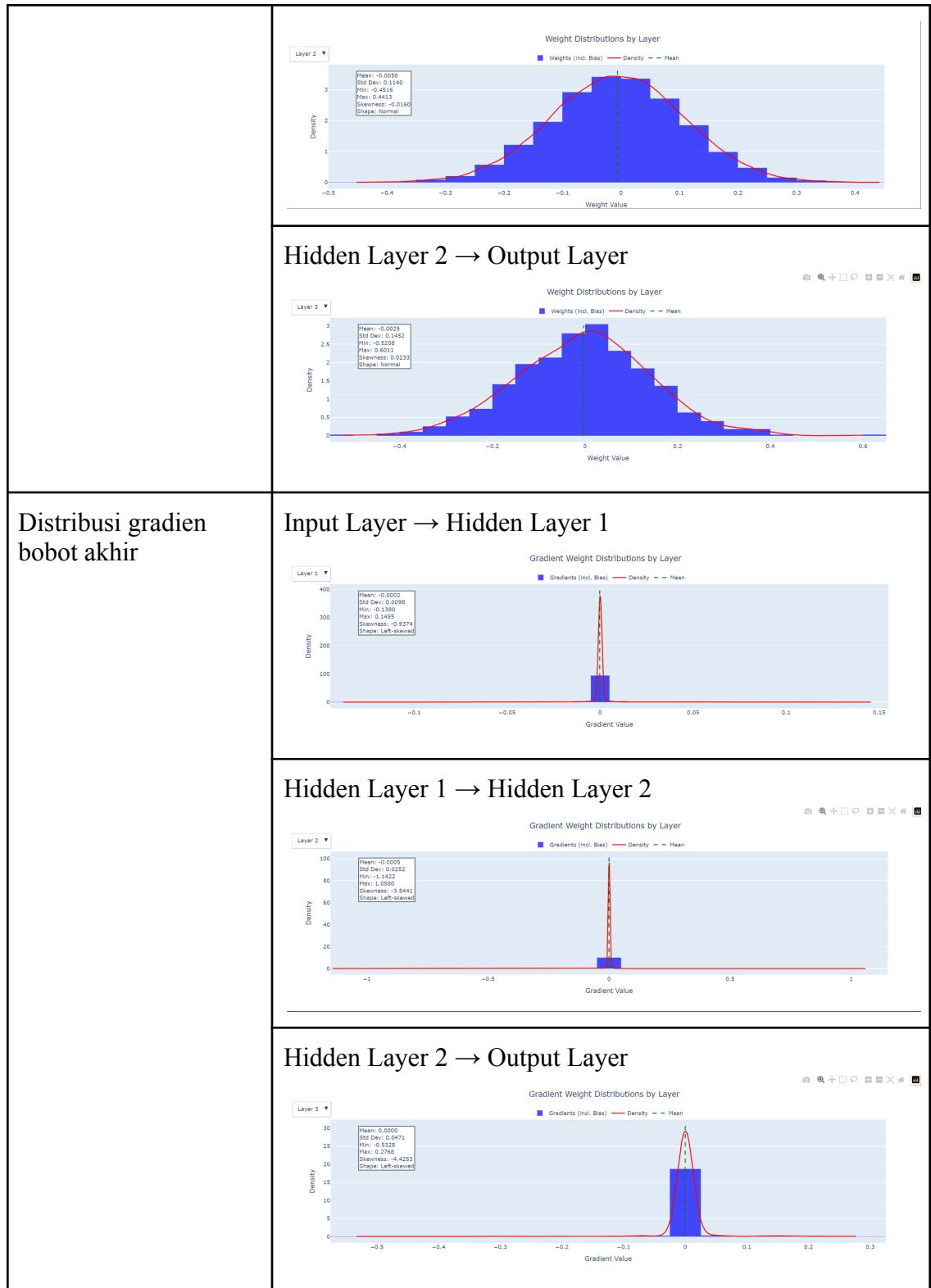


Distribusi bobot akhir

Input Layer → Hidden Layer 1



Hidden Layer 1 → Hidden Layer 2



BAB III

Kesimpulan dan Saran

3.1. Kesimpulan

Setelah melakukan berbagai macam pengujian pada beberapa parameter yang digunakan, didapatkan beberapa *insight* terkait pengaruh nilai parameter pada proses pelatihan model FFNN. Semakin besar nilai parameter *width* (lebar) yang ditandai dengan semakin banyaknya jumlah neuron dalam satu *hidden layer* memberikan performa yang lebih baik bagi hasil pelatihan model. Semakin besar nilai parameter *depth* (kedalaman) yang ditandai dengan semakin banyaknya jumlah *hidden layer* yang digunakan model memberikan performa yang lebih baik bagi hasil pelatihan model. Nilai parameter *learning rate* terbaik untuk performa suatu model memerlukan upaya tuning yang baik menyesuaikan parameter lain yang model gunakan untuk melakukan proses pelatihan. Pada parameter kontrol penggunaan 4 layer dengan jumlah neuron pada input layer sebanyak 784 neuron, hidden layer 1 sebanyak 150 neuron, hidden layer 2 sebanyak 75 neuron, dan output layer sebanyak 10 neuron, serta metode inisialisasi bobot Normal He, fungsi aktivasi ReLU untuk hidden layer, fungsi aktivasi softmax untuk output layer, fungsi loss Categorical Cross-Entropy, batch size sebanyak 64, dan jumlah epoch 20, learning rate terbaik yang dapat digunakan adalah 0.001.

Pemilihan metode inisialisasi bobot erat kaitannya dengan fungsi aktivasi yang digunakan. Diperlukan eksplorasi teoritis serta eksplorasi pelatihan model dari pemilihan metode inisialisasi serta fungsi aktivasi model untuk bisa mendapatkan hasil terbaik dari proses pelatihan model. Pada parameter kontrol penggunaan 4 layer dengan jumlah neuron pada input layer sebanyak 784 neuron, hidden layer 1 sebanyak 150 neuron, hidden layer 2 sebanyak 75 neuron, dan output layer sebanyak 10 neuron, fungsi loss Categorical Cross-Entropy, batch size sebanyak 64, dan jumlah epoch 20, didapatkan bahwa pada penggunaan fungsi aktivasi ReLU pada hidden layer, metode inisialisasi bobot He memberikan performa terbaik pada hasil pelatihan model.

Model FFNN yang penulis bangun ternilai sudah *robust* melihat kecilnya selisih nilai akurasi hasil pelatihan oleh FFNN menggunakan library sklearn dengan hasil pelatihan model FFNN buatan penulis pada penggunaan parameter kontrol yang sama. Penulis berhasil mengimplementasikan arsitektur FFNN pada model buatan yang sesuai dengan arsitektur FFNN ideal.

3.2. Saran

Untuk pengembangan lebih lanjut model, beberapa saran yang penulis berikan untuk mengoptimasi lebih lanjut hasil model yang dibangun adalah :

- Eksplorasi arsitektur model secara lebih mendalam dan lebih kompleks dengan memanfaatkan teknik regularisasi lainnya
- Eksplorasi tuning parameter yang lebih optimal
- Eksplorasi komparasi model buatan dengan model lainnya untuk mendapat perspektif yang lebih komprehensif terhadap pembuatan model untuk tugas klasifikasinya

LAMPIRAN

Pranala GitHub: https://github.com/salsbiila/Tubes1ML_G07_SSB

REFERENSI

K. Pykes, “Fighting Overfitting With L1 or L2 Regularization: Which One Is Better?,” neptune.ai, Apr. 04, 2021. <https://neptune.ai/blog/fighting-overfitting-with-l1-or-l2-regularization>