

# **“Laporan Analisis Klasifikasi Dataset Iris Menggunakan Neural Network dan Logistic Regression”**

**AI Kelompok 6 :**

Wafa Billah

Mauludha Fiozaki

Abbas Al Badawi

Sabri Mutiurrahman

Jauhan Ahmad

Muhammad Setya Adjie

May 13, 2025

## Pendahuluan

Dataset Iris adalah dataset klasik dalam machine learning yang berisi 150 sampel dari tiga spesies bunga Iris (Setosa, Versicolor, dan Virginica). Setiap sampel memiliki empat fitur: panjang sepal, lebar sepal, panjang petal, dan lebar petal, serta label spesies. Tujuan analisis ini adalah membangun model klasifikasi menggunakan Multi-Layer Perceptron (MLP) dan Logistic Regression untuk memprediksi spesies bunga berdasarkan fitur-fitur tersebut, serta membandingkan performa kedua model berdasarkan akurasi.

## Metodologi

### 1. Pengumpulan dan Persiapan Data

Berdasarkan hasil riset kami dataset ini mencakup 150 baris dan lima kolom: sepal\_length, sepal\_width, petal\_length, petal\_width, dan species. Langkah pra-pemrosesan meliputi:

- **Encoding Label:** Kolom species diubah menjadi numerik (0, 1, 2) menggunakan LabelEncoder.
- **Normalisasi Fitur:** Fitur numerik dinormalisasi menggunakan StandardScaler.
- **Pemisahan Data:** Data dibagi menjadi 80% data latih (120 sampel) dan 20% data uji (30 sampel) dengan train\_test\_split dan random\_state=42.
- **Konversi ke Tensor:** Input untuk MLP diubah menjadi tensor PyTorch.

### 2. Pembangunan Model

- **Neural Network (MLP):** Dibangun menggunakan PyTorch dengan arsitektur sebagai berikut:
  - Lapisan input: 4 neuron (sesuai fitur).
  - Lapisan tersembunyi: 10 neuron dengan aktivasi ReLU.
  - Lapisan output: 3 neuron (sesuai kelas).Model dilatih selama 100 epoch dengan optimizer Adam (learning rate 0.01) dan fungsi loss CrossEntropyLoss.
- **Logistic Regression:** Dibangun menggunakan scikit-learn dengan parameter max\_iter=200.

### 3. Evaluasi Model

Kedua model ini dievaluasi menggunakan akurasi (accuracy\_score) pada data uji. Proses pelatihan MLP divisualisasikan melalui plot loss terhadap epoch dengan hasil →

## Hasil

- **Neural Network (MLP):**
  - Loss menurun dari 0.8742 (epoch 10) menjadi 0.0845 (epoch 100).
  - Akurasi pada data uji: 100%.
  - Plot loss menunjukkan konvergensi yang mulus.

- **Logistic Regression:**
  - Akurasi pada data uji: 100%.
- Model MLP disimpan dalam file model.pth.

### Analisis

- **Performa Model:** Akurasi 100% pada kedua model menunjukkan dataset Iris memiliki pemisahan kelas yang jelas, terutama pada fitur petal\_length dan petal\_width.
- **Konvergensi MLP:** Penurunan loss yang konsisten mengindikasikan arsitektur MLP yang cukup untuk dataset ini.
- **Perbandingan Model:** Logistic Regression, meskipun sederhana, setara dengan MLP, hal ini menunjukkan dataset ini tidak memerlukan model kompleks.
- **Kelemahan:** Akurasi sempurna mungkin karena ukuran dataset ini kecil dan mudah. Evaluasi tambahan seperti confusion matrix atau cross-validation belum dilakukan.

### Kesimpulan

Dari hasil analisis ini kami berhasil membangun model MLP dan Logistic Regression dengan akurasi 100% pada data uji. Dataset Iris yang sederhana memungkinkan performa optimal, dengan input empat fitur yang efektif untuk klasifikasi. Namun, kami masih membutuhkan evaluasi lebih lanjut untuk memastikan generalisasi modelnya.