

Laporan Akhir Proyek Klasifikasi Dataset Iris

1. Pendahuluan

Dataset Iris adalah salah satu data yang sering dipakai buat belajar machine learning. Di dalamnya ada 150 data bunga iris dengan informasi panjang dan lebar kelopak (sepal) serta mahkota bunga (petal), lalu label spesiesnya (setosa, versicolor, virginica). Di proyek ini, kita mau coba dua pendekatan buat klasifikasi: Neural Network pakai PyTorch dan Logistic Regression pakai scikit-learn.

2. Tujuan Proyek

- Mengklasifikasi jenis bunga iris dari fitur panjang dan lebar sepal/petal
- Ngebangun model klasifikasi dengan neural network sederhana
- Bandingin hasilnya dengan Logistic Regression

3. Dataset

- Sumber data: Kaggle - Iris Flower Dataset
- Fitur: sepal_length, sepal_width, petal_length, petal_width
- Label: species (setosa, versicolor, virginica)

4. Pra-Pemrosesan

- Nama label diubah jadi angka biar bisa diproses model
- Semua fitur dinormalisasi biar model lebih gampang belajar (pakai MinMax Scaling)

5. Model 1: Neural Network (PyTorch)

Modelnya punya 1 hidden layer dengan 16 neuron, pakai fungsi aktivasi ReLU.

- Data dibagi jadi 80% buat training dan 20% buat testing
- Optimizer: Adam
- Loss function: CrossEntropyLoss
- Dilatih selama 50 epoch
- Akurasi di data test: sekitar 100%

Laporan Akhir Proyek Klasifikasi Dataset Iris

Contoh prediksi: [5.1, 3.5, 1.4, 0.2] diprediksi jadi Setosa

6. Model 2: Logistic Regression (scikit-learn)

- Data dibagi juga 80:20
- Normalisasi fitur tetap sama
- Model: LogisticRegression dari scikit-learn
- Akurasi test: sekitar 96.67%

Dari classification report, model ini udah cukup bagus buat semua kelas.

7. Perbandingan Model

- Neural Network lebih fleksibel dan akurat, tapi butuh lebih banyak resource
- Logistic Regression lebih simpel, gampang dimengerti, dan cepat dilatih
- Dua-duanya cocok dipakai tergantung kebutuhan

8. Kesimpulan

Hasil klasifikasi dari dua model ini udah sangat bagus. Neural Network cocok buat skenario yang lebih kompleks, sedangkan Logistic Regression cukup banget buat data yang sederhana kayak Iris.

9. Saran Pengembangan

- Coba model lain kayak SVM, KNN, atau Random Forest
- Tambahin validasi silang (cross-validation)
- Bisa juga dibikin versi GUI atau aplikasi web buat demo
- Tambahkan visualisasi biar datanya lebih gampang dimengerti

10. Referensi

- Dataset Iris dari UCI

Laporan Akhir Proyek Klasifikasi Dataset Iris

- Dokumentasi PyTorch
- Dokumentasi scikit-learn

Laporan Akhir Proyek Klasifikasi Dataset Iris

Lampiran: Rubrik Penilaian Tugas Kelompok

[illegible]