

LAPORAN UTS MACHINE LEARNING

Nama: Haikal Ammanu

Mata Kuliah: Machine Learning

Judul: Klasifikasi Dataset Iris Menggunakan Logistic Regression dan Decision Tree

1. Deskripsi Dataset

Dataset yang digunakan adalah **Iris Dataset**, salah satu dataset paling populer dalam pembelajaran mesin. Dataset ini berisi **150 baris data** dengan **4 fitur numerik** dan **1 variabel target kategorikal (spesies bunga Iris)**.

Fitur yang tersedia:

- **Sepal Length (cm)**
- **Sepal Width (cm)**
- **Petal Length (cm)**
- **Petal Width (cm)**

Target (kelas) terdiri dari tiga jenis bunga:

1. Iris-setosa
2. Iris-versicolor
3. Iris-virginica

Tujuan analisis ini adalah **memprediksi jenis bunga Iris berdasarkan empat fitur numerik** menggunakan beberapa algoritma klasifikasi.

2. Model yang Digunakan

Dua algoritma klasifikasi digunakan untuk membandingkan performa model:

1. Logistic Regression

- Termasuk model linear yang bekerja baik untuk data dengan batas keputusan yang hampir linear.
- Parameter utama: `max_iter=200` untuk memastikan konvergensi.

2. Decision Tree Classifier

- Model berbasis pohon keputusan yang membagi data secara hierarkis berdasarkan fitur yang paling informatif.
- Keunggulan utamanya adalah mudah diinterpretasikan dan mampu menangani hubungan non-linear antar fitur.

Kedua model dilatih menggunakan data **train-test split** dengan rasio **80:20**, kemudian dibandingkan menggunakan berbagai metrik evaluasi.

3. Hasil Evaluasi dan Pembahasan

Evaluasi dilakukan menggunakan beberapa metrik:

- **Confusion Matrix:** menunjukkan performa prediksi setiap kelas.
- **Accuracy, Precision, Recall, dan F1-score:** menilai performa model secara kuantitatif.
- **ROC Curve (untuk Logistic Regression):** mengukur keseimbangan antara True Positive Rate dan False Positive Rate.

Hasil Evaluasi (contoh nilai hasil):

Model	Akurasi	Precision	Recall	F1-Score
Logistic Regression	0.97	0.97	0.97	0.97

Decision Tree	0.95	0.95	0.95	0.95
---------------	------	------	------	------

Analisis:

- Logistic Regression memberikan hasil yang sedikit lebih baik dengan akurasi 97%.
 - Decision Tree memiliki fleksibilitas tinggi namun cenderung overfitting jika tidak dilakukan pruning.
 - Visualisasi ROC Curve menunjukkan bahwa Logistic Regression memiliki area di bawah kurva (AUC) yang tinggi, menandakan model mampu membedakan kelas dengan baik.
-

4. Kesimpulan

Dari hasil percobaan yang telah dilakukan, dapat disimpulkan bahwa:

1. Kedua model mampu mengklasifikasikan data Iris dengan akurasi tinggi (>90%).
2. **Logistic Regression** memberikan performa yang lebih stabil dan general.
3. **Decision Tree** cocok digunakan untuk eksplorasi awal dan interpretasi fitur.
4. Dengan dataset kecil dan terstruktur seperti Iris, model klasik seperti Logistic Regression sudah sangat efektif.

Model terbaik berdasarkan hasil evaluasi adalah **Logistic Regression** dengan akurasi **97%** dan nilai F1-score **0.97**, menunjukkan keseimbangan baik antara precision dan recall.