

**Nama Mahasiswa : Nurul Azizah Lonek**

**Nomor Induk Mahasiswa : 231011400457**

**Mata Kuliah : Machine Learning**

## **LAPORAN SINGKAT KLASIFIKASI DATASET IRIS**

### **1. Deskripsi Dataset**

Dataset yang digunakan dalam penelitian ini adalah dataset Iris, yang merupakan salah satu dataset paling populer dalam pembelajaran mesin. Dataset ini berisi 150 sampel bunga iris, dengan masing-masing memiliki empat fitur numerik yaitu:

<b>Fitur</b>	<b>Keterangan</b>
Sepal Length (cm)	Panjang kelopak bunga
Sepal Width (cm)	Lebar kelopak bunga
Petal Length (cm)	Panjang mahkota bunga
Petal Width (cm)	Lebar mahkota bunga

Target (label) memiliki tiga kelas:

- a. 0 : Iris Setosa
- b. 1 : Iris Versicolor
- c. 2 : Iris Virginica

Tujuan klasifikasi adalah untuk memprediksi jenis bunga iris berdasarkan keempat fitur di atas. Dataset ini bersifat seimbang (masing-masing kelas memiliki 50 data) dan tanpa nilai kosong, sehingga cocok digunakan untuk percobaan klasifikasi dasar.

### **2. Model yang Digunakan**

Pada tugas ini digunakan dua algoritma klasifikasi berbeda untuk membandingkan performanya, yaitu:

#### **a. Logistic Regression**

Logistic Regression adalah model linear classifier yang digunakan untuk memprediksi probabilitas suatu kelas berdasarkan fungsi logit. Model ini bekerja dengan baik pada data yang dapat dipisahkan secara linier.

Langkah-langkah:

1. Melakukan normalisasi data dengan StandardScaler
2. Melatih model Logistic Regression dengan data train
3. Melakukan prediksi pada data test
4. Mengevaluasi hasil dengan Confusion Matrix, Accuracy, Precision, Recall, F1-score, dan ROC Curve.

#### **b. Decision Tree**

Decision Tree adalah algoritma berbasis pohon yang membagi data berdasarkan aturan (if-then) pada setiap node. Model ini mudah diinterpretasikan dan dapat menangani data non-linear.

Langkah-langkah:

1. Membagi data train dan test tanpa normalisasi (karena tidak diperlukan pada Decision Tree)

2. Melatih model Decision Tree dengan data train
3. Melakukan prediksi pada data test
4. Mengevaluasi hasil dengan Confusion Matrix, Accuracy, Precision, Recall, dan F1-score.

### 3. Hasil Evaluasi dan Pembahasan

#### a. Logistic Regression

Confusion Matrix:

```
[[ 10  0  0]
 [  0  9  1]
 [  0  0 10]]
```

Accuracy	1.00
Precision	1.00
Recall	1.00
F1-score	1.00

ROC Curve

- AUC (Area Under Curve) untuk ketiga kelas berada di atas 1.00
- Menunjukkan kemampuan model yang sangat baik dalam membedakan antar kelas

#### b. Desition Tree

Confusion Matrix

```
[[10  0  0]
 [ 0 10  0]
 [ 0  0 10]]
```

Accuracy	1.00
Precision	1.00
Recall	1.00
F1-score	1.00

### 4. Perbandingan Hasil Antar Model

Model	Accuracy	Kelebihan	Kekurangan
Logistic Regression	100%	Cepat, stabil, mudah diinterpretasi	Kurang efektif untuk data non-linear
Decision Tree	100%	Akurasi tinggi, mudah dipahami	Berisiko overfitting, sensitif terhadap perubahan data

Analisis Perbandingan:

- Logistic Regression cocok digunakan untuk dataset besar dan linear.
- Decision Tree lebih baik untuk dataset kecil dan sederhana seperti Iris.
- Walaupun Decision Tree mencapai akurasi sempurna, Logistic Regression lebih kuat untuk generalisasi data baru.

## **5. Kesimpulan**

Berdasarkan hasil evaluasi kedua model:

1. Kedua model berhasil mengklasifikasikan dataset Iris dengan akurasi tinggi.
2. Decision Tree memiliki akurasi tertinggi (100%) dan tanpa kesalahan prediksi.
3. Logistic Regression tetap menunjukkan performa baik (100%) dan lebih stabil terhadap data baru.
4. Untuk dataset sederhana, Decision Tree lebih unggul, tetapi untuk dataset kompleks atau dengan banyak noise, Logistic Regression lebih direkomendasikan.

<https://github.com/NurulAzizahLonek26/Machine-Learning-uts-.git>