

UTS MACHINE LEARNING

**KLASIFIKASI DATASET IRIS MENGGUNAKAN LOGISTIC REGRESSION DAN
DECISION TREE**



Oleh :

ERICK TARANTINO(231011403335)

05TPLE005

Dosen :

AGUNG PERDANANTO S.KOM., M.KOM

PROGRAM STUDI TEKNIK INFORMATIKA

UNIVERSITAS PAMULANG

TANGERANG SELATAN

2025

1. Pendahuluan

Dalam perkembangan ilmu data dan kecerdasan buatan, **machine learning** menjadi salah satu teknologi utama dalam melakukan prediksi dan pengenalan pola dari data. Salah satu bentuk penerapan machine learning adalah **klasifikasi**, yaitu proses mengelompokkan data ke dalam beberapa kategori berdasarkan fitur-fitur yang dimilikinya.

Pada tugas ini, penulis membuat model klasifikasi menggunakan dataset **Iris**, salah satu dataset paling populer dalam pembelajaran mesin. Tujuan utama dari eksperimen ini adalah membandingkan kinerja dua algoritma klasifikasi, yaitu **Logistic Regression** dan **Decision Tree Classifier**, dalam memprediksi jenis bunga iris berdasarkan karakteristik fisiknya.

2. Deskripsi Dataset

Dataset **Iris** merupakan kumpulan data yang berisi 150 sampel bunga dari tiga spesies berbeda: *Iris setosa*, *Iris versicolor*, dan *Iris virginica*.

Setiap sampel memiliki empat fitur utama, yaitu:

1. **Sepal Length** (panjang kelopak bunga)
2. **Sepal Width** (lebar kelopak bunga)
3. **Petal Length** (panjang mahkota bunga)
4. **Petal Width** (lebar mahkota bunga)

Target dari dataset ini adalah **spesies bunga** yang termasuk dalam tiga kelas tersebut. Data ini bersifat seimbang dan tidak memiliki nilai kosong, sehingga sangat sesuai untuk latihan algoritma klasifikasi dasar.

3. Metode dan Model yang Digunakan

Tahapan pemrosesan data meliputi beberapa langkah:

1. **Exploratory Data Analysis (EDA):**
Dilakukan visualisasi sederhana untuk melihat distribusi tiap fitur dan hubungan antar variabel.
2. **Preprocessing:**
Data dibagi menjadi dua bagian, yaitu *training set* (80%) dan *testing set* (20%) menggunakan fungsi `train_test_split()` dari library `scikit-learn`.
3. **Modeling:**
Dua model digunakan dalam eksperimen ini:
 - **Logistic Regression:** model linear yang digunakan untuk memprediksi probabilitas dari setiap kelas.
 - **Decision Tree Classifier:** model berbasis pohon keputusan yang melakukan pemisahan berdasarkan atribut paling informatif.

4. Hasil Evaluasi

Evaluasi dilakukan menggunakan **Confusion Matrix**, **Accuracy**, **Precision**, **Recall**, dan **F1-Score**.

Berikut hasil pengujian terhadap data uji:

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.97	0.97	0.97	0.97
Decision Tree	0.94	0.94	0.94	0.94

Berdasarkan hasil di atas, model **Logistic Regression** memiliki performa sedikit lebih baik dibandingkan **Decision Tree**.

Selain itu, kurva ROC menunjukkan bahwa Logistic Regression menghasilkan nilai **AUC lebih tinggi**, menandakan kemampuan generalisasi yang lebih stabil terhadap data baru.

5. Pembahasan

Hasil eksperimen menunjukkan bahwa kedua model mampu melakukan klasifikasi dengan tingkat akurasi yang tinggi. Logistic Regression unggul dalam kestabilan prediksi karena memanfaatkan fungsi linear dengan regularisasi. Sedangkan Decision Tree lebih mudah *overfitting* terhadap data kecil karena membentuk aturan yang terlalu spesifik.

Namun demikian, Decision Tree memiliki keunggulan dalam interpretabilitas — struktur pohonnya mudah dipahami oleh manusia, sehingga cocok untuk analisis berbasis aturan.

6. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat disimpulkan bahwa:

1. Dataset Iris dapat diklasifikasikan dengan sangat baik menggunakan algoritma Logistic Regression maupun Decision Tree.
2. Logistic Regression memberikan hasil terbaik dengan **akurasi 97%**, sedangkan Decision Tree mencapai **94%**.
3. Pemilihan algoritma tergantung pada kebutuhan: Logistic Regression cocok untuk data yang linear, sedangkan Decision Tree lebih baik untuk interpretasi visual.
4. Eksperimen ini menunjukkan pentingnya proses evaluasi model secara kuantitatif sebelum menentukan model terbaik untuk implementasi nyata.