

LAPORAN TUGAS UTS

MODEL KLASIFIKASI MENGGUNAKAN DATASET IRIS

Nama: Aditya Eka Purnama

NIM: 231011400988

Kelas: 05TPLE004

Program Studi: Teknik Informatika

Dosen Pengampu: Agung Perdananto, S.Kom., M.Kom.

1. Pendahuluan

Tugas ini bertujuan untuk membuat model klasifikasi menggunakan dataset dengan target kategorikal.

Pada laporan ini, digunakan dataset *Iris* yang berisi data karakteristik bunga iris untuk memprediksi jenis spesiesnya. Proses pembuatan model dilakukan dengan tahapan pemilihan dataset, eksplorasi data (EDA), pembersihan dan preprocessing data, pembuatan model dengan dua algoritma, serta evaluasi hasil prediksi.

2. Deskripsi Dataset

Dataset *Iris* merupakan salah satu dataset paling populer dalam pembelajaran mesin.

Dataset ini terdiri

dari 150 baris data dengan empat fitur utama, yaitu panjang dan lebar sepal (sepal length, sepal width),

serta panjang dan lebar petal (petal length, petal width).

Target atau variabel kelas terdiri dari tiga jenis bunga iris:

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

3. Model dan Metode

Dalam eksperimen ini digunakan dua algoritma klasifikasi yaitu:

1. Logistic Regression
2. Decision Tree Classifier

Langkah-langkah pengerjaan meliputi:

- Membagi dataset menjadi data latih dan data uji dengan rasio 80:20.
- Melakukan normalisasi fitur menggunakan *StandardScaler*.
- Melatih model dengan data latih dan menguji menggunakan data uji.
- Mengevaluasi hasil menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *ROC Curve*.

4. Hasil Evaluasi

Setelah dilakukan pelatihan dan pengujian terhadap model, diperoleh hasil evaluasi sebagai berikut:

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.97	0.97	0.97	0.97
Decision Tree	0.93	0.93	0.93	0.93

Dari tabel di atas, dapat dilihat bahwa model Logistic Regression memberikan performa yang sedikit lebih baik dibandingkan Decision Tree pada dataset Iris. Hal ini menunjukkan bahwa Logistic Regression lebih stabil terhadap data yang memiliki distribusi linier.

5. Pembahasan

Kedua model mampu mengklasifikasikan data dengan cukup baik karena dataset Iris relatif sederhana dan terpisah dengan baik.

Decision Tree cenderung lebih fleksibel, namun memiliki risiko *overfitting*. Logistic Regression memberikan hasil yang lebih stabil karena menggunakan pendekatan linier terhadap fitur.

ROC Curve yang dihasilkan menunjukkan area di bawah kurva (AUC) yang tinggi untuk kedua model, dengan Logistic Regression memiliki nilai AUC sedikit lebih besar, menunjukkan kemampuan klasifikasi yang lebih baik.

6. Kesimpulan

Berdasarkan hasil pengujian dan evaluasi, dapat disimpulkan bahwa:

1. Logistic Regression memberikan performa terbaik pada dataset Iris dengan nilai akurasi 97%.
2. Decision Tree juga memberikan hasil baik, namun sedikit lebih rendah.
3. Proses klasifikasi berjalan efektif dan sesuai dengan tujuan pembelajaran model supervised learning.

Dengan demikian, Logistic Regression dapat direkomendasikan sebagai model terbaik untuk klasifikasi dataset Iris.