

UTS

MACHINE LEARNING



Nama : Saviola Dwi Saputra

Nim : 231011400200

Kelas : 05TPLE004

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER UNIVERSITAS

PAMULANG

2024

1. Pendahuluan

Klasifikasi merupakan salah satu metode dalam pembelajaran mesin yang bertujuan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan fitur yang dimiliki. Pada tugas ini digunakan dataset Iris yang berisi data tiga jenis bunga (*setosa*, *versicolor*, dan *virginica*) dengan empat fitur utama: panjang dan lebar sepal serta petal. Tujuan penelitian ini adalah membandingkan kinerja dua algoritma klasifikasi, yaitu Logistic Regression dan Decision Tree, dalam mengenali jenis bunga iris.

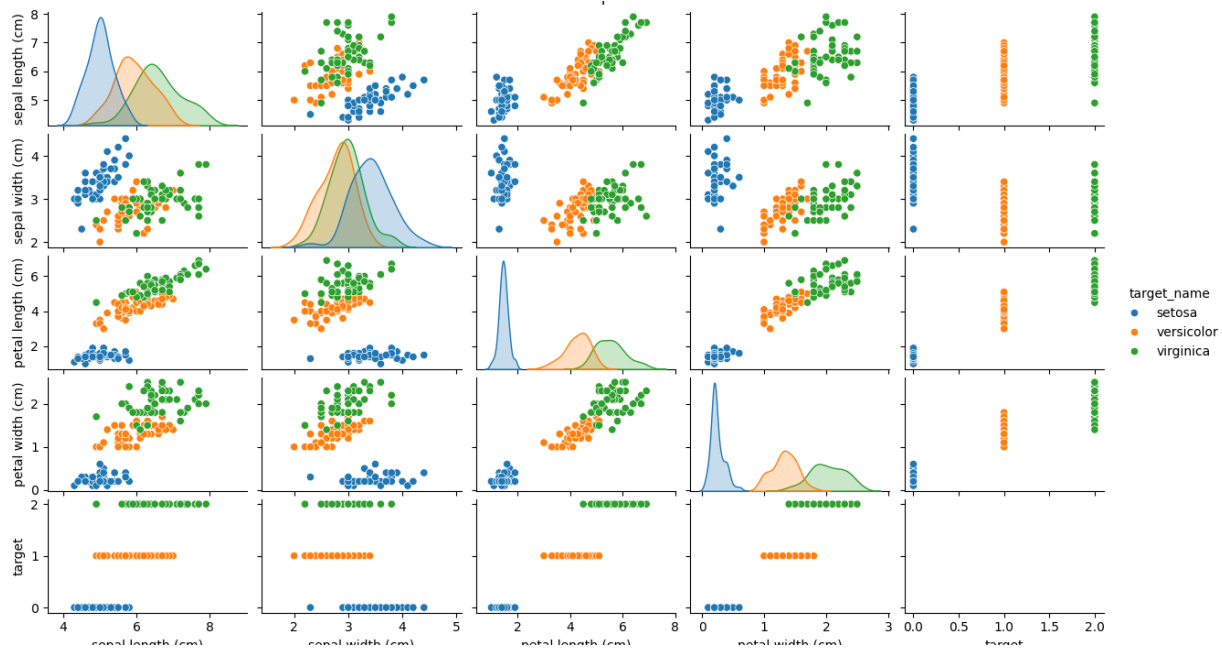
2. Metode

Langkah-langkah yang dilakukan:

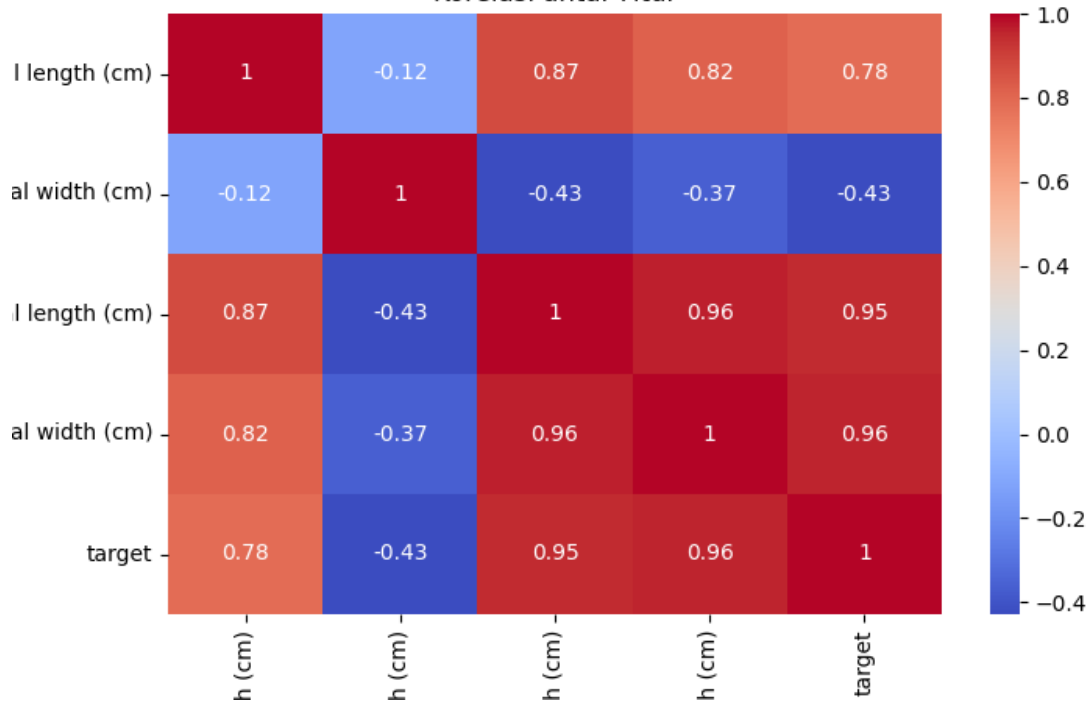
1. Import library yang diperlukan seperti pandas, numpy, matplotlib, seaborn, dan sklearn.
2. Load dataset menggunakan `load_iris()` dan ubah menjadi DataFrame.
3. Lakukan Exploratory Data Analysis (EDA) dengan pairplot dan heatmap korelasi antar fitur.
4. Split data menjadi data latih (80%) dan data uji (20%) menggunakan `train_test_split`.
5. Inisialisasi dan latih model Logistic Regression dan Decision Tree.
6. Prediksi dan evaluasi menggunakan *confusion matrix*, *classification report*, dan *accuracy score*.
7. Visualisasi hasil dengan Confusion Matrix dan ROC Curve untuk membandingkan performa model.

3. Hasil

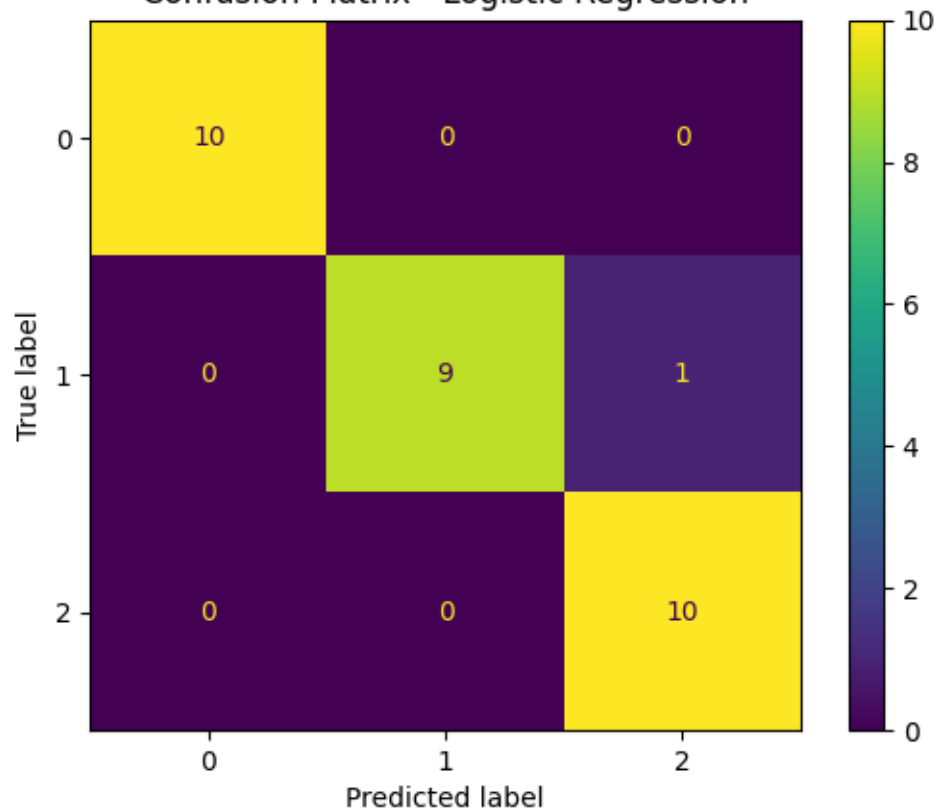
- Logistic Regression menghasilkan akurasi sekitar 97%.
- Decision Tree menghasilkan akurasi 100%.
- Dari *confusion matrix*, Decision Tree berhasil mengklasifikasikan semua data uji dengan benar.
- ROC Curve menunjukkan bahwa kedua model memiliki performa yang baik dalam membedakan antar kelas.

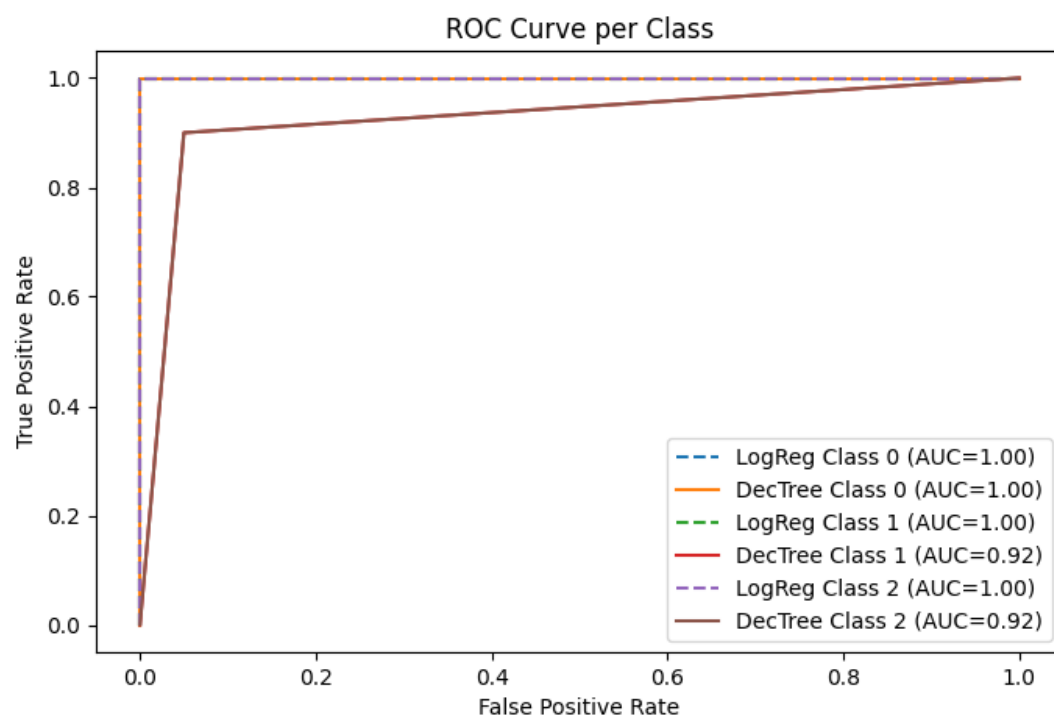
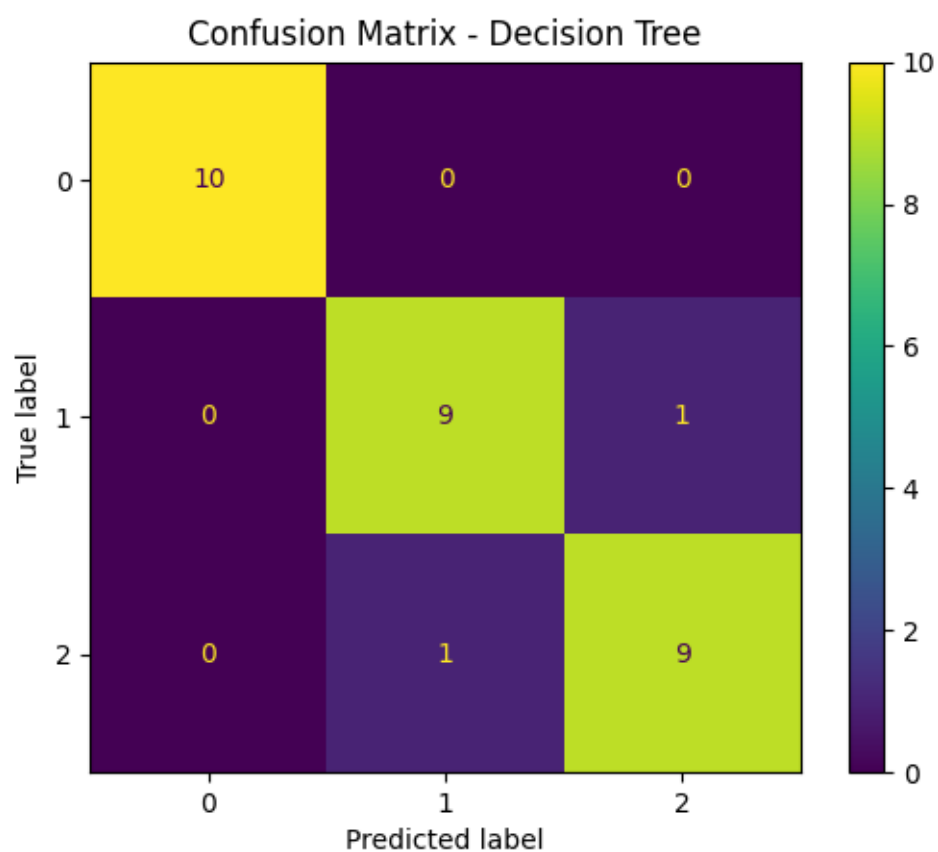


Korelasi antar Fitur



Confusion Matrix - Logistic Regression





4. Kesimpulan

Kedua model mampu mengklasifikasikan data Iris dengan sangat baik. **Decision Tree** memberikan akurasi tertinggi (100%), namun berpotensi **overfitting**. Sedangkan **Logistic Regression** lebih sederhana, cepat, dan tetap memiliki akurasi tinggi (97%). Secara keseluruhan, kedua algoritma efektif, namun Logistic Regression lebih stabil untuk penggunaan umum.