

Technical Report

Giovanni Nathaniel

1103202211

Kode yang disediakan adalah skrip Python yang melakukan analisis data eksplorasi (EDA), pelatihan model, dan evaluasi pada dataset kanker payudara. Berikut adalah langkah-langkah utama yang dilakukan dalam skrip:

1. Impor perpustakaan yang diperlukan seperti NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, dan Graphviz.
2. Muat dataset kanker payudara menggunakan fungsi `load_breast_cancer()` dari Scikit-learn dan ubah menjadi Pandas DataFrame.
3. Pisahkan data menjadi set pelatihan dan pengujian menggunakan fungsi `train_test_split()` dari Scikit-learn.
4. Visualisasikan data menggunakan dua plot estimasi kepadatan kernel (KDE) dan peta panas menggunakan pustaka Seaborn.
5. Latih pengklasifikasi pohon keputusan pada set pelatihan asli, set fitur yang dikurangi, dan set fitur yang diubah PCA menggunakan fungsi `DecisionTreeClassifier()` dari Scikit-learn.
6. Mengevaluasi keakuratan pengklasifikasi terlatih pada set pelatihan dan pengujian.
7. Visualisasikan classifier pohon keputusan terlatih menggunakan fungsi `export_graphviz()` dari library Graphviz.
8. Plot kepentingan fitur menggunakan fungsi `plot_feature_importances()`.
9. Plot kurva belajar menggunakan fungsi `plot_learning_curve()`.

Kesimpulannya, skrip memberikan contoh yang baik tentang cara melakukan EDA, pelatihan model, dan evaluasi menggunakan Scikit-learn dan Seaborn. Kode dapat ditingkatkan lebih lanjut dengan menambahkan lebih banyak komentar dan memberikan penjelasan terperinci untuk setiap langkah.