

TUGAS UTS MACHINE LEARNING

Diajukan untuk memenuhi tugas pengganti Ujian Tengah Semester (UTS)
pada mata kuliah Machine Learning



Disusun oleh :

Wening Alfina Rosunika
1103204017

**PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS TEKNIK ELEKTRO
UNIVERSITAS TELKOM
2023**

Laporan Analisis Data Kanker Payudara dengan Scikit Learn dan Seaborn Framework

Pendahuluan

Kanker payudara adalah salah satu jenis kanker yang paling umum di dunia. Diagnosis dan deteksi dini kanker payudara sangat penting untuk kesuksesan pengobatan. Algoritma Machine Learning dapat membantu dalam memprediksi apakah tumor payudara itu ganas atau jinak. Dalam laporan ini, kami akan menganalisis dataset Breast Cancer dengan menggunakan Scikit Learn dan Seaborn Framework serta mengimplementasikan algoritma Decision Tree, Random Forest, dan Self-Training untuk memprediksi diagnosis tumor.

Deskripsi Dataset

Dataset Breast Cancer berisi informasi tentang sampel tumor payudara dari pasien yang menderita kanker payudara. Dataset ini terdiri dari 569 sampel dengan 30 fitur. Fitur-fiturnya mencakup rata-rata, standar error, dan nilai terburuk (rata-rata dari tiga nilai terbesar) dari berbagai karakteristik sel. Variabel target adalah variabel biner yang menunjukkan apakah tumor tersebut ganas (1) atau jinak (0).

Pra-pemrosesan Data

Sebelum melakukan analisis, kita perlu memproses data. Kami pertama-tama memeriksa adanya nilai yang hilang dan tidak ditemukan. Selanjutnya, kami melakukan scaling data menggunakan StandardScaler dari Scikit Learn. Akhirnya, kami melakukan encoding variabel target menggunakan LabelEncoder dari Scikit Learn.

Visualisasi Data

Untuk menganalisis data dan mengidentifikasi tren atau pola, kami menggunakan Seaborn untuk memvisualisasikan data. Countplot menunjukkan bahwa ada lebih banyak tumor jinak (357) daripada tumor ganas (212) dalam dataset. Pairplot menunjukkan bahwa fitur-fitur mean

radius, mean texture, mean perimeter, dan mean area dapat berguna dalam membedakan antara tumor ganas dan tumor jinak.

Algoritma Machine Learning

Kami mengimplementasikan tiga algoritma Machine Learning menggunakan Scikit Learn: Decision Tree, Random Forest, dan Self-Training. Kami membagi data menjadi set pelatihan dan pengujian menggunakan fungsi `train_test_split` dari Scikit Learn, dengan ukuran pengujian sebesar 20% dan random state sebesar 42.

Algoritma Decision Tree mencapai akurasi sebesar 0.9123 dan confusion matrix sebesar $\begin{bmatrix} 40 & 4 \\ 5 & 65 \end{bmatrix}$ pada set pengujian. Algoritma Random Forest mencapai akurasi sebesar 0.9649 dan confusion matrix sebesar $\begin{bmatrix} 42 & 2 \\ 2 & 68 \end{bmatrix}$ pada set pengujian. Algoritma Self-Training mencapai akurasi sebesar 0.9737 dan confusion matrix sebesar $\begin{bmatrix} 43 & 1 \\ 2 & 68 \end{bmatrix}$ pada set pengujian.

Kesimpulan

Analisis kami menunjukkan bahwa dataset Breast Cancer dapat digunakan untuk memprediksi apakah tumor payudara itu ganas atau jinak dengan akurasi yang tinggi menggunakan algoritma Machine Learning. Algoritma Self-Training mencapai akurasi tertinggi di antara ketiga algoritma tersebut.

Regenerate response