

Machine Learning Technical Report

“Breast Dataset Cancer Clasification Using Decision Tree,Random Forest and Self-Training ”



**Universitas
Telkom**

Oleh:

Muhammad Haidar Abdul Jabbar

1103202071

Program Studi Teknik Komputer

Fakultas Teknik Elektro

Universitas Telkom

2022/2023

Pendahuluan

Deteksi dini kanker payudara merupakan hal yang penting dalam mengatasi masalah kesehatan yang serius. Decision Tree, Random Forest, dan Self-Training adalah beberapa algoritma machine learning yang dapat digunakan dalam mengklasifikasikan kanker payudara berdasarkan dataset massa payudara. Metode ini memungkinkan untuk mengembangkan model prediksi yang dapat membantu dalam memprediksi jenis tumor dan memperkirakan kemungkinan metastasis kanker payudara. Hasil dari penggunaan model ini diharapkan dapat meningkatkan deteksi dini dan pengobatan kanker payudara, sehingga dapat mengurangi angka kematian akibat kanker payudara.

Dataset

Dalam Implementasinya, digunakannya Dataset Breast Cancer Wisconsin (Diagnostic) untuk mengklasifikasikan kanker payudara dengan menggunakan tiga algoritma pembelajaran mesin: Decision Tree, Random Forest, dan Self-Training. Dataset ini terdiri dari 569 sampel, dimana setiap sampel memiliki 30 fitur yang diambil dari gambar digital aspirasi jarum halus (FNA) dari massa payudara.

Visualisasi

Dalam visualisasi data menggunakan seaborn yang merupakan salah satu library python yang didesain untuk memperindah visualisasi data dan memudahkan analisis statistik pada data. Terutama dalam implementasi kode digunakan salah satu jenis visualisasi PairPlot

Data Preprocessing

Untuk mempersiapkan Dataset Kanker Payudara Wisconsin (Diagnostik) untuk algoritma Machine Learning, langkah preprocessing dilakukan. Dataset dipartisi menjadi set pelatihan dan pengujian, dengan rasio pembagian 70:30. Fitur-fiturnya juga distandarisasi menggunakan fungsi StandardScaler dari Library scikit-learn, yang diperlukan untuk memastikan bahwa semua fitur berada dalam skala yang sama dan dapat meningkatkan kinerja algoritma pembelajaran mesin.

Algoritma Klasifikasi(Decision Tree, Random Forest)

Model Decision Tree, Random Forest, dan Self Training dilatih pada dataset yang telah diproses sebelumnya menggunakan DecisionTreeClassifier, RandomForestClassifier, ataupun SelfTrainingClassifier dari library scikit-learn. Model dievaluasi pada set pengujian dan mencapai akurasi yang mendekati nilai mendekati sempurna.

Kesimpulan

Perbandingan kerja tiga algoritme Machine Learning untuk klasifikasi kanker payudara menggunakan Dataset Kanker Payudara Wisconsin. Algoritme Random Forest mengungguli algoritma Decision Tree dan Self-Training dalam hal akurasi, presisi, daya ingat, dan skor F1. Seaborn juga digunakan untuk memvisualisasikan hubungan antara fitur-fitur dalam kumpulan data. Yang dimana hasil ini dapat digunakan untuk meningkatkan akurasi deteksi dan pengobatan kanker payudara.