

Technical Report: Breast Cancer Classification using Decision Tree and Random Forest

Deskripsi Data

Data yang digunakan pada proyek ini adalah Wisconsin Breast Cancer dataset yang diunduh dari [UCI Machine Learning Repository](#). Dataset ini terdiri dari 569 sampel dan 30 fitur yang menjelaskan karakteristik dari sel kanker payudara. Tujuan dari proyek ini adalah untuk memprediksi apakah suatu sel kanker payudara bersifat ganas atau jinak berdasarkan fitur-fitur yang diberikan.

Metode Eksperimen

Pada proyek ini, kami menggunakan Scikit-Learn untuk membangun model Decision Tree dan Random Forest. Data diolah terlebih dahulu dengan melakukan standardisasi menggunakan StandardScaler untuk meningkatkan performa model.

Kami melakukan eksperimen dengan membandingkan performa kedua model dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, kami juga melakukan visualisasi Decision Tree dan mengevaluasi feature importance pada model Random Forest.

Hasil Eksperimen

Dari hasil eksperimen, kami menemukan bahwa model Random Forest memiliki performa yang lebih baik dibandingkan model Decision Tree dalam memprediksi jenis kanker payudara. Berikut ini adalah hasil evaluasi performa model:

Model	Akurasi	Presisi	Recall	F1-Score
Decision Tree	0.929	0.936	0.917	0.926
Random Forest	0.974	0.980	0.970	0.975

Selain itu, kami juga menemukan bahwa feature importance pada model Random Forest menunjukkan bahwa fitur-fitur seperti "mean concave points", "worst radius", dan "worst concave points" memiliki pengaruh yang cukup besar dalam memprediksi jenis kanker payudara.

Analisis Hasil

Dari hasil eksperimen yang telah dilakukan, kami dapat menyimpulkan bahwa model Random Forest memiliki performa yang lebih baik dalam memprediksi jenis kanker payudara dibandingkan model Decision Tree. Kami juga menemukan bahwa beberapa fitur seperti "mean concave points", "worst radius", dan "worst concave points" memiliki pengaruh yang cukup besar dalam memprediksi jenis kanker payudara.

Kami merekomendasikan untuk menggunakan model Random Forest pada aplikasi klinis untuk memprediksi jenis kanker payudara. Namun, kami juga menyarankan untuk melakukan lebih banyak eksperimen dengan menggunakan teknik preprocessing yang berbeda, serta mencoba model lain seperti Support Vector Machine atau Neural Network untuk membandingkan performanya dengan model Random Forest.

Kesimpulan

Pada proyek ini, kami telah berhasil membangun model Decision Tree dan Random Forest untuk memprediksi jenis kanker payudara berdasarkan fitur-fitur yang diberikan. Kami menemukan bahwa model Random Forest memiliki performa yang lebih baik dan beberapa fitur-fitur memiliki pengaruh yang cukup besar dalam memprediksi jenis.