

Technical Report
Machine Learning Model Using Python



Telkom
University

Oleh:

Nama : Zaidan Luthfi

NIM : 1103203238

PROGRAM STUDI TEKNIK KOMPUTER

FAKULTAS TEKNIK ELEKTRO

UNIVERSITAS TELKOM

2023

I. Pendahuluan

Kanker payudara adalah kondisi di mana sel-sel abnormal di dalam jaringan payudara berkembang secara tidak terkendali dan membentuk tumor. Tumor ini dapat menjadi ganas atau jinak, tetapi kanker payudara merujuk pada tumor yang bersifat ganas. Kanker payudara adalah salah satu jenis kanker yang paling umum di kalangan perempuan, meskipun juga dapat terjadi pada laki-laki. Tanda-tanda dan gejala kanker payudara meliputi benjolan pada payudara, perubahan pada bentuk atau ukuran payudara, perubahan pada kulit payudara atau puting susu, keluar cairan dari puting susu, dan pembengkakan kelenjar getah bening di ketiak.

Pembuatan machine learning menggunakan dataset breast cancer dari scikit merupakan salah satu contoh implementasi teknologi yang digunakan untuk mendukung pengambilan keputusan di bidang kesehatan. Dataset ini adalah salah satu dataset kanker payudara yang paling populer digunakan untuk melatih model machine learning.

Dataset breast cancer dari scikit dibangun oleh tim peneliti kesehatan yang terdiri dari Dr. William H. Wolberg, Dr. W. Nick Street, dan Olvi L. Mangasarian. Dataset ini berisi informasi tentang tumor payudara dari 569 pasien perempuan dengan 30 fitur yang diukur dari gambar citra digital. Pemanfaatan dataset breast cancer dari scikit dalam pembuatan model machine learning bertujuan untuk membantu memprediksi apakah tumor tersebut bersifat jinak atau ganas. Model machine learning dapat mempelajari pola dari data yang ada untuk membangun algoritma yang dapat digunakan untuk mengklasifikasikan tumor payudara menjadi kategori yang sesuai.

II. Persiapan dan Pengolahan Data

Program ini menggunakan beberapa model machine learning seperti Regresi Logistik (Self training), Random Forest, dan Pohon Keputusan, serta beberapa teknik pengolahan data seperti normalisasi menggunakan StandardScaler dan pembagian data menjadi set pelatihan dan pengujian.

Pertama-tama, dataset diubah menjadi DataFrame dan ditampilkan lima baris pertama dari dataset. Kemudian, dilakukan visualisasi distribusi variabel target dan matriks korelasi menggunakan seaborn. Setelah itu, dataset dibagi menjadi set pelatihan dan pengujian dengan proporsi pengujian sebesar 30% menggunakan train_test_split dan dilakukan normalisasi menggunakan StandardScaler.

Selanjutnya, dilakukan pelatihan tiga model machine learning, yaitu regresi logistic self training, random forest, dan pohon keputusan menggunakan set pelatihan yang telah di-normalisasi. Kemudian, ketiga model tersebut dievaluasi pada set pengujian dan akurasi masing-masing model dicetak di layar.

Selain akurasi, juga dilakukan visualisasi hasil evaluasi pada set pengujian menggunakan classification report dan confusion matrix. Classification report menampilkan nilai precision, recall, dan f1-score pada masing-masing kelas, sedangkan confusion matrix menampilkan nilai true positive, false positive, false negative, dan true negative pada masing-masing model.

Selanjutnya, dilakukan visualisasi tingkat kepentingan fitur pada model random forest menggunakan barplot. Terakhir, dilakukan visualisasi dari model pohon keputusan menggunakan plot_tree untuk memvisualisasikan struktur pohon keputusan.

III. Kesimpulan

Program ini menggunakan beberapa model machine learning seperti Regresi Logistik (Self training), Random Forest, dan Pohon Keputusan, serta beberapa teknik pengolahan data seperti normalisasi menggunakan StandardScaler dan pembagian data menjadi set pelatihan dan pengujian. Pada hasil running dari ketiga model tersebut, model self-training menempati akurasi tertinggi sebesar 98.24%, diikuti oleh model random forest sebesar 97.07% dan model decision tree mempunyai akurasi sebesar 94.15%.

Model-model AI, seperti Random Forest, Decision Tree, dan Self Training, dapat digunakan untuk membantu dokter dalam deteksi kanker payudara dengan tingkat akurasi yang cukup tinggi. Model-model ini dapat mengambil data dari berbagai faktor seperti usia, ukuran tumor, jumlah sel tumor, dan lainnya untuk memberikan hasil yang dapat membantu dokter dalam membuat keputusan yang lebih tepat.

Alasan mengapa model-model tersebut dapat membantu dokter dalam deteksi kanker payudara adalah karena mereka dapat melakukan analisis dan pengolahan data secara cepat dan akurat. Hal ini dapat membantu dokter untuk memperoleh informasi yang lebih akurat dan cepat dalam memutuskan diagnosa dan perawatan yang tepat untuk pasien.

Selain itu, model-model AI juga dapat membantu dalam mendeteksi kanker payudara pada tahap awal sehingga dapat meningkatkan peluang kesembuhan pasien. Dengan bantuan teknologi AI, dokter dapat memperoleh informasi yang lebih tepat dan akurat dalam memutuskan tindakan selanjutnya yang harus dilakukan terhadap pasien.