

DETEKSI BREAST CANCER DENGAN PENGEMBANGAN MACHINE LEARNING MODEL

Studi Kasus : Bojonegoro



NITA CAHYANI, ALIF YUANITA KARTINI,
IVA NUR ACHYANI, HIMMATUL LUTVIANA

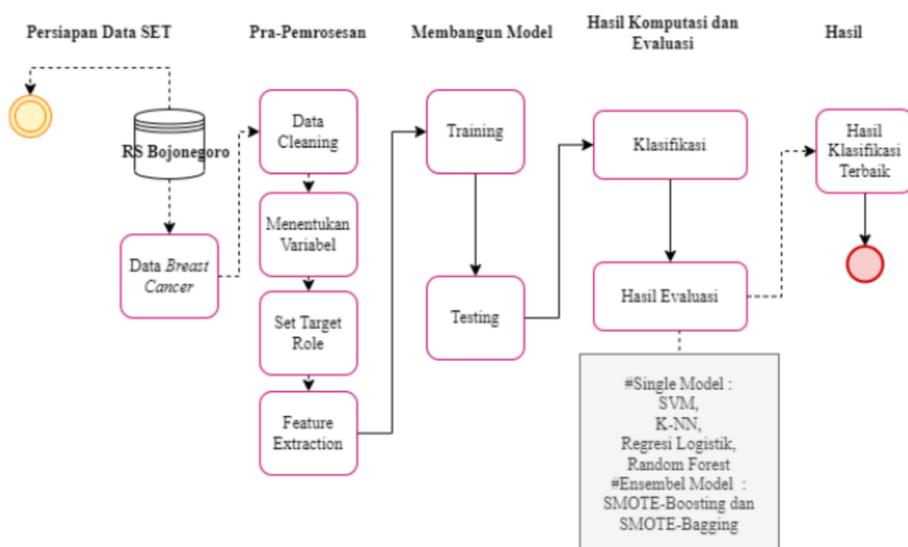
LATAR BELAKANG

Breast Cancer atau Kanker payudara adalah penyakit yang paling umum ditemukan pada wanita di seluruh dunia. Setiap perkembangan untuk prediksi dan diagnosis penyakit kanker merupakan modal penting untuk hidup sehat. Sehingga dibutuhkan prediksi secara akurat. Dalam penelitian ini, diterapkan algoritma Machine Learning (aplikasi dari Artificial Intelligence) yaitu metode single: Support Vector Machine (SVM), Random Forest, Logistic Regression, dan K-Nearest Neighbors (KNN) dan metode ensemble yaitu SMOTE-Boosting dan SMOTE-Bagging

TUJUAN

Mendapatkan ketepatan klasifikasi atau prediksi breast cancer khususnya studi kasus di Bojonegoro dengan tingkat kinerja yang lebih baik.

METODE

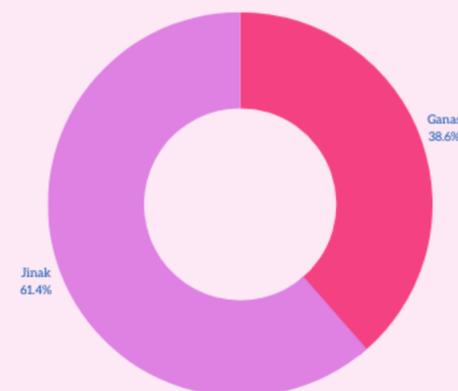


Gambar 1 Arsitektur Penelitian

Variabel penelitian yang digunakan dalam penelitian ini, yaitu 2 kelas (jinak dan ganas), dan 8 atribut (usia, jumlah jaringan, berat (grm), ukuran (cm3), inti sel, kromatin, mitosis, penyebaran). Gambar 1 Arsitektur dalam penelitian.

HASIL

Berdasarkan Gambar 2 diketahui bahwa proporsi kanker ganas sebesar 38,56% sedangkan proporsi kanker jinak sebesar 61,44%, IR (Imbalance Ratio) adalah ukuran yang paling umum digunakan untuk menggambarkan tingkat ketidakseimbangan set data. Ketika $IR = 1$, artinya dataset yang benar-benar seimbang. Ketika $IR > 1$, semakin besar IR, semakin besar tingkat ketidakseimbangan dataset. Dalam penelitian ini IR sebesar 1.59%.



Gambar 2 Persentase Status Breast Cancer

Gambar 3 diketahui nilai akurasi terbesar data training maupun data testing yaitu 100% diperoleh dengan menggunakan metode ensemble SMOTE-Boosting dengan algoritma pembelajaran yang digunakan adalah random forest dengan menggunakan tingkat pengambilan sampel berlebih dari SMOTE sebesar 300, ukuran ensemble yang digunakan sebesar 20, Jumlah k diterapkan dalam algoritma SMOTE adalah 5, Jumlah pohon keputusan (decision trees) di setiap forest model ensemble adalah 100.



Gambar 3 Persentase Nilai Akurasi Training & Testing

Berdasarkan Gambar 3 dapat disimpulkan bahwa metode ensemble SMOTE-Boosting RF merupakan metode yang terbaik dalam penelitian ini.