

METODE

Penelitian ini mengembangkan model prediktif untuk memperkirakan risiko kekambuhan kanker tiroid menggunakan tiga algoritma *machine learning*: *Random Forest*, *XGBoost*, dan *Logistic Regression*. Tabel 1 merupakan data penelitian diperoleh dari rekam medis dan sampel sel pasien RSUD DR Ken Saras setelah mendapatkan persetujuan etik, mencakup 14 variabel klinis dan demografis. Variabel target adalah diagnosis (jinak/ganas), sementara variabel prediktor meliputi usia, jenis kelamin, riwayat keluarga, paparan radiasi, status yodium, kebiasaan merokok, obesitas, diabetes, kadar hormon tiroid (TSH, T3, T4), serta ukuran nodul tiroid.

Tabel 1 Variabel Penelitian

No	Nama Variabel
1	Umur
2	Jenis kelamin
3	<i>Family_History</i>
4	<i>Radiation_Exposure</i>
5	Kekurangan Yodium
6	Merokok
7	Obesitas
8	Diabetes
9	Tingkat Hormon Perangsang Tiroid
10	Tingkat triiodotironin (ng/dL).
11	Tingkat tiroksin ($\mu\text{g/dL}$).
12	Ukuran nodul tiroid (cm)
13	Perkiraan risiko kanker tiroid
14	Diagnosis

Logistic Regression dipilih sebagai model baseline karena kesederhanaannya dan kemampuannya memberikan interpretasi klinis melalui *odds ratio*. Model ini memetakan hubungan linear antara prediktor dan outcome menggunakan fungsi logit dengan optimasi maximum likelihood estimation. *XGBoost* diimplementasikan untuk menangkap hubungan non-linear dan interaksi kompleks melalui pendekatan *gradient boosting* dengan mekanisme pruning dan paralelisasi. Model ini memerlukan tuning hiperparameter (*learning rate*, *tree depth*, dll) untuk mencegah *overfitting*. *Random Forest* digunakan sebagai model ensemble berbasis bagging yang membangun banyak *decision tree* secara paralel, dengan keunggulan mengurangi variance prediksi dan robust terhadap *noise data*.

Evaluasi model dilakukan secara komprehensif menggunakan metrik *akurasi*, *precision*, *recall*, dan *AUC-ROC*. Validasi menggunakan *stratified fold cross-validation* dan *hold-out validation* pada data independen untuk memastikan generalisasi model. Pendekatan ini memungkinkan pemilihan model yang tidak hanya unggul secara statistik tetapi juga relevan klinis, dengan mempertimbangkan keseimbangan antara sensitivitas deteksi dan spesifisitas diagnosis dalam konteks prediksi kekambuhan kanker tiroid.