

Abstrak

Kanker paru-paru merupakan penyebab kematian paling umum pada seseorang yang menderita kanker. Hal ini terjadi karena mengingat pentingnya fungsi paru-paru sebagai alat pernapasan dan distribusi oksigen ke seluruh tubuh. Identifikasi dini kanker paru-paru sangat penting untuk mengurangi angka kematiannya. Akurasi sangat penting karena menunjukkan seberapa akurat model atau sistem membuat prediksi yang tepat. Tingkat akurasi yang tinggi menunjukkan bahwa model tersebut dapat menghasilkan temuan yang dapat dipercaya dan akurat, yang penting untuk membuat keputusan yang efektif berdasarkan data yang tersedia. Dalam penelitian ini digunakan pendekatan pembelajaran ensemble yaitu metode bagging dan boosting untuk mengklasifikasikan kanker paru. Hyperparameter, suatu kelas parameter, sangat penting untuk efektivitas model ini. Untuk meningkatkan akurasi model klasifikasi kanker paru-paru, penyelidikan menyeluruh dilakukan untuk mengidentifikasi kombinasi hyperparameter terbaik. Pada penelitian ini, dataset yang digunakan adalah dataset medis yang berisi riwayat pasien yang terdiagnosis kanker paru atau tidak. Dataset diambil dari Kaggle mysarahmadbhat dan cancerdatahp dari data world. Untuk mengevaluasi keakuratan model, penelitian ini menggunakan metode matriks konfusi yang membandingkan hasil prediksi model dengan kebenaran dasar. temuan penelitian mengungkapkan bahwa penggunaan rasio pemisahan kumpulan data 70:30 memberikan hasil terbaik, dengan model Random Forest, CatBoost, dan XGBoost mencapai akurasi 98%, presisi 0,98, perolehan 0,98, dan skor f1 0,98. namun untuk AdaBoost, hasil terbaik diperoleh pada dataset dengan rasio 80:20 dengan akurasi 96%, presisi 0,97, recall 0,96, dan f1-score 0,96. Mengoptimalkan pemilihan parameter yang tepat dan menggunakan pembagian data latih yang proporsional dapat meningkatkan efektivitas model dalam mengklasifikasi kanker paru-paru.