

ABSTRAK

Kanker paru-paru masih menjadi salah satu penyebab utama kematian akibat kanker di seluruh dunia, sehingga diperlukan upaya diagnosis yang lebih dini dan akurat. Penelitian ini bertujuan untuk mengevaluasi peran pemilihan fitur dalam meningkatkan kinerja model pembelajaran mesin untuk diagnosis kanker paru-paru. Dengan memanfaatkan *Lung Cancer Dataset*, metode pemilihan fitur berbasis peringkat seperti *Information Gain* dan *Gini Decrease* digunakan untuk menentukan fitur-fitur paling penting dalam proses pelatihan model. Tiga algoritma klasifikasi yang digunakan dalam penelitian ini meliputi *Random Forest*, *Decision Tree*, dan *Support Vector Machine (SVM)*. Kinerja masing-masing model dievaluasi melalui sejumlah metrik, yaitu AUC, akurasi klasifikasi, F1 score, presisi, recall, dan MCC. Hasil penelitian menunjukkan bahwa penerapan pemilihan fitur memberikan peningkatan yang signifikan terhadap kinerja model, serta berkontribusi dalam mengurangi kompleksitas komputasi dan meningkatkan ketepatan diagnosis. Model *Random Forest* menunjukkan performa terbaik dengan nilai AUC sebesar 0,980 dan akurasi sebesar 94,4%, yang mengindikasikan tingkat keandalan yang tinggi dalam memprediksi hasil diagnosis kanker paru-paru. Temuan ini menegaskan pentingnya integrasi pemilihan fitur dalam alur kerja pembelajaran mesin guna membangun model yang efisien, dapat diinterpretasikan, dan aplikatif dalam praktik klinis.

Kata Kunci: Pemilihan Fitur, Diagnosis Kanker Paru-paru, Machine Learning, Random Forest