

Model Klasifikasi berbasis Ekspresi Gen *Non-Small Cell Lung Carcinoma* (NSCLC) pada Wanita Bukan Perokok Menggunakan Metode *Ensemble*

Kholishoh Nur Azizah, Fhira Nhita, Isman Kurniawan,

Fakultas Informatika, Universitas Telkom, Bandung
kholishoh@students.telkomuniversity.ac.id, fhiranhita@telkomuniversity.ac.id,
ismankrn@telkomuniversity.ac.id

Abstrak

Kanker paru-paru adalah penyebab utama kematian terkait kanker di seluruh dunia dan membawa dampak sosial ekonomi yang signifikan bagi pasien, keluarga, dan masyarakat secara keseluruhan. Dalam diagnosis kanker, klasifikasi berbagai jenis tumor sangat penting. Prediksi akurat dari berbagai jenis tumor memungkinkan untuk pengobatan yang lebih baik dan meminimalkan toksisitas pada pasien. Untuk menganalisis masalah klasifikasi kanker menggunakan data ekspresi gen, untuk pemilihan fitur dan model prediksi. Penelitian ini bertujuan untuk memprediksi NSCLC dengan menerapkan metode *ensemble* pada data *microarray*. Penulis menggunakan tiga metode *ensemble* untuk memprediksi NSCLC, yaitu *Random Forest*, *Adaptive Boosting (AB)*, dan *Extreme Gradient Boosting (XG)*. Seleksi fitur dilakukan menggunakan *variance threshold* dan parameter *chi-square* kemudian dilanjutkan dengan membangun model prediksi menggunakan *ensemble*. Hasil validasi model terbaik berdasarkan yang terdiagnosis kanker yaitu model AB dengan 10 fitur, XG dengan 10 fitur, dan XG dengan 20 fitur yang menghasilkan nilai *accuracy*, *recall*, dan *f1-score* yang sama berturut-turut yaitu 0.93, 1.00, dan 0.93.

Kata kunci: *chi-square*, *ensemble*, *microarray*, *variance threshold*