

Fitur Seleksi pada Data Microarray untuk Deteksi Kanker Berdasarkan Klasifikasi Random Forest

Tita Nurul Nuklianggraita¹, Adiwijaya², Annisa Aditsania³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹titanggraita@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstrak

Kanker adalah penyakit yang dapat menyerang semua organ manusia. Berdasarkan data *World Health Organization* (WHO) fact-sheets tahun 2018, kematian akibat kanker telah mencapai 9.6 juta jiwa. Salah satu cara yang dikenal untuk mendeteksi kanker yaitu Teknik *Microarray*, tetapi data *microarray* memiliki dimensi yang besar karena jumlah fitur yang sangat banyak dibandingkan dengan jumlah sampelnya. Oleh karena itu, reduksi dimensi harus dilakukan untuk menghasilkan akurasi yang optimal. Dalam makalah ini, membandingkan *Minimum Redundancy Maximum Relevance* (MRMR) dan *Least Absolute Shrinkage and Selection Operator* (LASSO) untuk mengurangi dimensi pada data *microarray*. Selain itu, dengan menggunakan Klasifikasi *Random Forest* (RF), kinerja klasifikasi (pada deteksi kanker) juga dibandingkan. Berdasarkan simulasi, dapat disimpulkan bahwa LASSO lebih baik daripada MRMR karena dapat menghasilkan evaluasi 100% pada lung cancer dan ovarian cancer, 92% pada colon cancer, 93% pada prostate tumor dan 83% pada central nervous system.

Kata kunci : Kanker, *Microarray*, *Minimum Redundancy Maximum Relevance* (MRMR), *Least Absolute Shrinkage and Selection Operator* (LASSO), *Random Forest*.