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

Tita Nurul Nuklianggraita[1], Adiwijaya2[2], Annisa Aditsania3[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]titanggraita@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]aaditsania@telkomuniversity.ac.id

Abstract
Cancer is a disease that can affect all organs of humans. Based on data from the World Health Organization (WHO) fact sheet in 2018, cancer deaths have reached 9.6 million. One known way to detect cancer that is with Microarray Technique, but the microarray data have large dimensions due to the number of features that are very much compared to the number of samples. Therefore, dimension reduction should be made to produce optimum accuracy. In this paper, we compare Minimum Redundancy Maximum Relevance (MRMR) and Least Absolute Shrinkage and Selection Operator (LASSO) to reduce dimension of microarray data. Moreover, by using Random Forest (RF) Classifier, the performance of classification (cancer detection) is compared. Based on simulation, it can be concluded that LASSO is better than MRMR because it can produce an evaluation of 100% in lung and ovarian cancer, 92% colon cancer, 93% prostate tumor and 83% central nervous system.

Keywords – Cancer, Microarray, Minimum Redundancy Maximum Relevance (MRMR), Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest.