Deteksi Kanker pada Data *Microarray* Menggunakan Metode *Naïve Bayes* dengan *Hybrid Feature Selection*

Bintang Peryoga[1], Adiwijaya[2], Widi Astuti[3]

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

Abstract

Cancer is a deadly disease that is responsible for 9.6 million death in 2018 based on WHO data so early cancer detection is needed so can be treated immediately and cancer deaths can be reduced. Microarray is technology that can monitor and analyze the expression of cancer genes in microarray data but has high data dimension and small sample so dimensional reductions are needed for the optimal classification process. Dimension reduction can reduce the use of features for the classification process by selecting some influential features. Hybrid method is one dimension reduction by combining Filter method with Wrapper so it gets the both advantage. In this case, researchers combined Naïve Bayes with Hybrid Feature Selection (Information Gain - Genetic Algorithm) on cancer data for microarray Lung Cancer, Ovarian Cancer, Breast Cancer, Colon Tumors, and Prostate Tumors. These data were obtained from Kent-Ridge Biomedical Dataset. The results showed that from 5 data used, 4 data obtained an accuracy between 87-100% while the prostate tumor data obtained the smallest accuracy of 61.14%. The implementation of the feature selection method and the classification of the 5 cancer data above only uses less than 63 features to obtain this accuracy.

Keywords: cancer, microarray, naïve bayes, information gain, genetic algorithm, hybrid feature selection.