Abstrac

Cancer is one of the deadliest diseases in the world. Every year, cancer patients continue to increase and cause many casualties. Until now, the cure for this deadly disease is still hard to find. In recent years, microarray data technology has been widely used to diagnose cancer from an early age, DNA microarray data is a technology used to see sequences of nucleic acid sequences located at specific locations in DNA structures that can be used to analyze thousands of samples at the same time so that later can be classified which are classified as cancer and not cancer. Therefore, microarray data is data that has very large data dimensions. Data whose dimensions are very large can result in the calculation results being not optimal and the resulting classification accuracy is small. To optimize the data and increase the classification accuracy value of the large dimension data, dimensional reduction was carried out by selecting the Genetic Algorithm (GA) feature. Genetic Algorithm is usually able to provide good results and a fairly good level of accuracy. Classification of microarray data using the Naive Bayes method and Logistic Regression. The best accuracy of Genetic Algorithm and Logistic Regression classification in colon tumor data and mll leukemia are 100% and 57,7778% for Genetic Algorithm and Naive Bayes classification. And 67% accuracy get from Logistic Regression in the mll leukemia data.

Keywords : cancer, feature selection, classification, microarray data, genetic algorithm, naive bayes, logistic regression