Abstract

In an advanced era such as the current detection of cancer can be done in several ways, one of which is bioinformatics by using microarray technology. The technology consists of DNA that forms microchips with very large dimensions. Large size dimensions cause computational calculations. To reduce computational problems, the reduction is done before being classified using (Tree Classification and Regression) CART. Dimension reduction by selecting components, this component is selected because not all attributes in the microarray data are selected, considering that the data on the microarray is very large. Components that have the most characteristics are chosen so that calculations can produce optimal results. Dimension reduction used in this study is feature extraction using the principal component analysis (PCA) principle. Feature extraction is usually used for continuous data by extracting attributes so that they can produce attributes. There are three cancer data used, namely, colon cancer, leukemia, and lung cancer. The accuracy generated from this study averages over 70% with the PCA algorithm for reducing dimensions and CART as its classification.

Key word: cancer, microarray, dimention reduction, CART