Abstract

Cancer is one of many deadly diseases. But cancer is not the source of the real disease because the fact is every adult every day there will be 30 billion cells that die and 1 to 10 percent of them are cancer cells, The point is every person must have cancer cells. Cancer grows in any part of the body, one of them is the lungs. Based on data from the World Health Organization (WHO) lung cancer is one type of cancer that affects many men, it is reinforced by the International Agency for Research on Cancer (IARC) that in 2012 in Indonesia there are 25,322 cases of lung cancer in men and 9374 cases in women. Lung cancer also has various types, this study tried to classify data on gene expression of lung cancer. Gene expression data in this study are complex, then dimension reduction or feature selection are required and the method used is Principal Components Analysis (PCA). For the algorithm for classification used Naive Bayes and also Support Vector Machine (SVM). In this study, the two classification algorithms are working together and will be examined how optimal it is. The accuracy results from the research using classification with Naive Bayes and SVM is 100%.