

ABSTRACT

A system that able to analyze and identify a person whether infected by a disease will extremely helping in medical world, because the increasingly death rates each day. One of the factors of death is the like of early treatment for the patients that are already infected by disease. This happens because the patients aren't realized that they're infected by deadly diseases. There are five diseases that are categorized as deadly which are cancer, heart attack, diabetes, AIDS, and TBC.

By that reason, on this final project will be built a system that able to predict whether a patient is infected by a disease or not. This system will help on the early treatment for patients that are caught by diseases. Data of cancer, heart attack, diabetes, AIDS, and TBC diseases a sourced from *Kent Ridge Bio-medical Data Set Repository* website will be used to build this system, in which these data are high dimensional data for each disease. These data has thousands of attributes that will be divided into two data, which are *training* data and *testing* data. Both data will be reduced by genetic algorithm (GA) method and classified by naïve bayes classifier method.

By predicting using this model, the result will show whether a patient have disease or not. Next is accuracy test using *testing* data to obtain the valid accuracy result. Therefore, from the result it shows that crossvalidation method is better with 88.89% accuracy for *colon tumor* data and 100% accuracy for *leukemia* data, compared to precentage split method with 78.95% accuracy for *colon tumor* data and 77.27% accuracy for *leukemia* data.

Keyword: Genetic Algorithm, Naïve Bayes, Evolutionary Data Mining