

ABSTRACT

In the data processing in data mining, searching information from high dimensional data becomes more difficult. The phenomenon about Curse of Dimensionality is the real problems related to high-dimensional data. That phenomenon may makes the process of data processing becomes less effective and efficient, so it requires a certain method to reduce the dimensions of the data.

The merger between data mining algorithms with Evolutionary Algorithms became a solution to overcome the problems associated with dimensional curse phenomenon. K-Nearest Neighbor (KNN) algorithm is a data mining algorithm that can be used to classify data and GA can help maximize the classification accuracy subset of attributes.

In this final projects will be built a system to predict the disease with a high dimension of data. The data that used for the system is Colon Tumor disease and leukemia. The best performance is obtained when the parameter P_c 0.6 and 0.1 P_m with a performance of 93.57% on Colon Tumor data and 100% on Leukemia data.

Keyword: high dimensional data, evolutionary data mining, k-nearest neighbor, genetic algorithm