

ABSTRACT

The more advanced technology that allows researchers in conducting the extraction level of gene expression in different types of experiments. This makes it easy to learn more about gene function, gene networks, biological processes and effects of a medical treatment. The first step to do this is to perform gene clustering. We have had a lot of clustering algorithms that are made but very few clustering algorithm that is specific to the gene data clustering. CLICK or Cluster Identification via Connectivity Kernels is one of algorithms that aims to perform data clustering of genes.

The algorithm is based on a weighted graph and use the calculation of probability that is expected to provide more accurate results. In this thesis performed the implementation and performance analysis of algorithms CLICK in performing data clustering of genes that are homogeneous or heterogeneous which will then be tested through the method of Figure Of Merit (FOM), as well as to find out whether the factor of a homogeneous or heterogeneous nature of the data will affect the value of FOM from these data. Also in this study the FOM value of the CLICK algorithm will be compared with FOM value generated by K-Means, GenClust and Cast algorithm in doing the same data clustering. Making it easier for the researchers in selecting the algorithm that will be used to perform data clustering of genes that have these character.

From the results of research conducted a conclusion that for the same amount of experiment, CLICK algorithm has a better performance in doing data clustering of genes that are more heterogeneous than homogeneous data when compared with the three other algorithms. While the FOM value is not affected by the characters of a heterogeneous or homogeneous data but by the average of the resulting clusters and the number of experimentally tested.

Keywords : CLICK, clustering, genes