

ABSTRACT

Data DNA microarray or gene expression is high-dimensional data describes the process regarding the unravelling of DNA to produce proteins. DNA microarray or gene expression is used to predict how the expression of every gene. Gene expression very much its benefits one of which is to find out the type of progression of the disease in the human body. The research of gene expression data had previously done on one journal that has published the "Review on Feature Selection of Gene Expression Data for Autism Classification" [8].

The study describes the gene expression data using a Genetic Algorithm. On this final assignment problem solved was of gene expression data using K-means Clustering algorithm used as classification in determining the grade of a sample of data and methods of the WPCA are used as selected features to reduce the dimensions of each attribute so that the number of attributes will be even simpler to process to the next stage. As for the feature selection using the right function to apply the weights on the WPCA method i.e. able to give quite a good influence against each attribute that has reduced the performance result is obtained that is 78,94% for number of 37 attributes with 80% training, 20% testing data proportions and 71,05% for number of 16 attributes with 20% testing, 80% training data proportions

Keywords: *Gene Expressions, WPCA, K-means Clustering*