ABSTRACT

With the widespread computerization in business, government, and science, the efficient and effective discovery of interesting patterns from large databases becomes essential. Data mining emerges as a solution to the data analysis probems faced by many organization. One of data mining functionality is *clustering* that is grouping data into *cluster*s depends on their similarities.

Subspace clustering is development in the *clustering* method, which finds *cluster*s in a dataset by selecting the most relevant dimensions for each *cluster* separately. FINDIT finds *cluster*s with *subspace clustering* based on two key ideas: *dimension-oriented distance* measure which fully utilizes dimensional difference information, and *dimension voting* policy. This final project has been implemented FINDIT algorithm and analysed the performance consider amount of data, dimension size of level to time and also consider *Dmindist* parameter of resultant *cluster*s accuracy.

User parameter *Dmindist* influence performance of software. Small or to over the value of *Dmindist*, resultant *cluster* accuracy become low, with missing one or more subspace in original *cluster* at the process. Increasing amount of data and dimension size will cause more time to get the result.

Keywords : data mining, *subspce clustering*, FINDIT *algorithm*, *dimension oriented distance*, *dimension voting*, *Dmindist*