ABSTRACT

K-MEANS ALGORITHM OPTIMIZATION USING *DIFFERENTIAL EVOLUTION* IN DETERMINING *CLUSTER CENTROID* POINTS

By

Muhammad Abdul Latief

21110002

The K-Means *algorithm is widely used for data clustering because of its straightforwardness and efficiency. Nevertheless, it has notable limitations. A primary issue is the randomness in selecting initial centroids, leading to inconsistent outcomes. Given that data clustering is essentially a global optimization problem, the choice of an objective function is crucial. Traditionally,* K-Means *employs the Euclidean distance function for its objective, but this can sometimes yield non-distinct clusters. To alleviate this issue, the concept of Least Square is introduced. Minimizing the objective function, based on the centroid, necessitates an appropriate method. This study examines the use of metaheuristic approaches (such as* Differential Evolution*), the integration of* Differential Evolution *with* K-Means*, and formulates a Least Square objective function to optimize centroids. During the computational phase, indicators like* Silhouette Score *and execution time are tracked. The results show that the combination of Differential Evolution (DE) and K-Means is more efficient based on the indicators used. In addition, this experiment also compares the performance between K-Means+DE and K-Means+GA (Genetic Algorithm) to evaluate the advantage of DE over other population-based metaheuristic algorithms. The analysis results show that K-Means+DE is superior in many experiments.To test the quality of this model, an implementation is carried out on the original data, namely the 2021 Village Potential Survey data obtained from the Central Bureau of Statistics. The implementation results show a high silhouette value, which is around 0.81884 at the number of clusters of 2.*

Keywords: Computational Performance; **Differential Evolution*; Initial Centroid;* K-Means*; Optimization.***