

ABSTRAK

IMPLEMENTASI ALGORITMA GENETIKA DALAM PENENTUAN *CENTROID* AWAL *K-MEANS* PADA PENGELOMPOKAN DATA PENYAKIT *TUBERKULOSIS*

Gita Shavira Pratiwi

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Trunojoyo Madura

K-Means adalah algoritma pengelompokan *non-hierarkis* yang paling sering digunakan karena kemampuannya untuk menangani data dalam jumlah besar. Namun, kelemahan utama *K-Means* adalah hasil pengelompokan yang sangat dipengaruhi oleh inisialisasi *centroid* yang dilakukan secara acak. Akibatnya, hasil pengelompokan sering kali terjebak pada optimum lokal. Sehingga membutuhkan semacam mekanisme untuk mencapai optimal global dengan menentukan inisialisasi *centroid* yang tepat. Salah satu cara paling efektif untuk mengatasi masalah *K-Means* adalah dengan menggunakan Algoritma Genetika. Dengan mengoptimasi titik pusat awal, yang pada akhirnya akan menjadi input *K-Means*. Sehingga titik pusat dioptimalkan menggunakan Algoritma Genetika, metode ini menghasilkan kinerja yang lebih optimal dibandingkan dengan metode *K-Means* biasa, yaitu iterasi yang lebih cepat dibandingkan dengan perhitungan *K-Means* tanpa optimasi dan pendekatan *GAK-Means* menghasilkan evaluasi yang lebih tinggi berdasarkan hasil pengujian yang telah dilakukan. Berdasarkan hasil uji coba parameter, diperoleh parameter yang paling optimal yaitu pada populasi 100, generasi 80, serta kombinasi $Cr = 0.4$ dan $Mr = 0.1$ dengan rata-rata tertinggi nilai *Silhouette* sebesar 0.754 pada $k = 5$, dan pada *K-Means* rata-rata nilai *Silhouette* tertinggi yang dihasilkan sebesar 0.698 pada $k = 3$.

Kata kunci: *Tuberkulosis, Clustering, Algoritma K-Means, Algoritma Genetika.*

ABSTRACT

IMPLEMENTATION OF GENETIC ALGORITHM IN DETERMINING INITIAL CENTROID K-MEANS FOR CLUSTERING OF TUBERCULOSIS DISEASES DATA

Gita Shavira Pratiwi

Department of Informatics Engineering, Faculty of Engineering

Trunojoyo University Madura

K-Means is the most frequently used non-hierarchical clustering algorithm because of its ability to handle large amounts of data. However, the main drawback of K-Means is that the clustering results are heavily influenced by random initialization of centroids. As a result, clustering results are often stuck at local optimum. So it requires some kind of mechanism to achieve global optimal by determining the proper initialization centroid. One of the most effective ways to solve the K-Means problem is to use a Genetic Algorithm. By optimizing the initial center point, which will eventually become K-Means input. So that the center point is optimized using the Genetic Algorithm, this method produces more optimal performance compared to the usual K-Means method, which is a faster iteration compared to K-Means calculations without optimization and the GAK-Means approach produces a higher evaluation based on the test results. have been done. Based on the results of the parameter trials, the most optimal parameters were obtained, namely in a population of 100, 80 generations, and the combination of $Cr = 0.4$ and $Mr = 0.1$ with the highest average Silhouette value of 0.754 at $k = 5$, and on average K-Means the highest Silhouette value produced was 0.698 at $k = 3$.

Keywords: Tuberculosis, Clustering, K-Means Algorithm, Genetic Algorithm.