

BAB I

PENDAHULUAN

A. Latar Belakang

Clustering merupakan proses pengelompokan data menjadi beberapa kelompok. Data yang dikelompokkan dalam *cluster* memiliki tingkat kemiripan yang tinggi dan data lintas *cluster* memiliki kemiripan yang rendah. *Clustering* adalah metode pengambilan data yang tidak terkontrol. Pendekatan tidak terkontrol yaitu suatu metode yang dilakukan tanpa pelatihan dan tidak memerlukan hasil. (Setiawan, Herwindiati and Sutrisno, 2019). Pada *clustering* terdapat pencarian global. Pencarian global menghasilkan pusat cluster lebih optimal karena tidak mudah terjebak dalam optimasi local. Salah satu metode untuk *clustering* yang bersifat pencarian global adalah Algoritma Genetika (AG). Menurut (Maiti and Bidinger, 2019) AG merupakan proses pencarian yang *metaheuristic* sehingga penekanan pemilihan operator yang digunakan sangat menentukan keberhasilan AG dalam menemukan solusi optimum suatu masalah.

Menurut (Setiawan, Herwindiati and Sutrisno, 2019) Algoritma Genetika untuk *clustering* menggunakan operator *roulette wheel selection* dan *arithmetic crossover* serta data citra bunga dengan skenario berbeda menghasilkan kinerja yang baik. Menurut (Akay, Tekeli and Yüksel, 2020) AG dengan *new fitness* menggunakan dataset sebagai data uji menghasilkan *clustering* yang lebih baik dibandingkan beberapa algoritma yang lain

Pengujian merupakan factor penting untuk mengetahui validitas dari hasil penelitian. Hasil pengujian memberikan informasi seberapa baik *clustering* dataset yang telah dilakukan (Abdullah, Winarko and Musdholifah, 2020). Salah satu metode evaluasi *clustering* yang digunakan untuk melihat seberapa optimal dari suatu *cluster* adalah *silhouette coefficient*. *Silhouette coefficient* mengevaluasi hasil *clustering* dengan menghitung nilai *silhouette index* dari sebuah data ke-*i* dibutuhkan untuk menghitung nilai *silhouette coefficient*. Nilai *silhouette coefficient* diperoleh dengan mencari nilai maksimal dari nilai *Silhouette Index Global* dari jumlah

cluster sampai dengan jumlah *cluster* $n-1$. Pada penelitian yang dilakukan oleh (Kuntjoro, Setiawan and Perdana, 2018) menggunakan *silhouette coefficient* untuk mengevaluasi hasil *clustering* menggunakan *K-means* dan AG. Hasil evaluasi menggunakan AG dan *K-means* berdasarkan *silhouette coefficient* yaitu 0.9959 dan *K-means* biasa 0.8831. Penelitian yang dilakukan (Fitriani, Musdholifah and Hartati, 2018) menggunakan *silhouette coefficient* untuk mengevaluasi hasil *clustering* menggunakan *Adaptive Unified Differential Evolution* (AuDE). Nilai hasil evaluasi *clustering* berdasarkan *silhouette coefficient* yaitu 1.

Berdasarkan tinjauan literature penelitian sebelumnya, maka dilakukan penelitian untuk pengembangan Algoritma Genetika dengan *new fitness* yang dikombinasikan dengan *roulette wheel selection* dan *arithmetic crossover* untuk *clustering* data, karena belum diketahui hasil evaluasi atau permancennya. Oleh karena itu, penelitian ini berjudul " **Evaluasi Hasil Clustering Menggunakan Algoritma Genetika dengan New Fitness, Roulette Wheel Selection dan Arithmetic Crossover** " untuk mengecek seberapa baik nilai evaluasi hasil *clustering* menggunakan AG dengan *new fitness*, *roulette wheel selection* dan *arithmetic crossover* berdasarkan *silhouette coefficient*.

B. Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka rumusan masalah dari penelitian ini adalah bagaimana mengukur nilai evaluasi hasil *clustering* menggunakan Algoritma Genetika dengan *new fitness*, *roulette wheel selection* dan *arithmetic crossover* berdasarkan *silhouette coefficient*?

C. Batasan Masalah

Batasan masalah dari penelitian ini yaitu sebagai berikut:

1. Menggunakan 5 dataset sebagai data uji yaitu dataset Iris, Glass, Wine, Ecoli, dan WDBC.

2. Evaluasi yang digunakan yaitu *silhouette coefficient*.

D. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui nilai evaluasi hasil *clustering* menggunakan Algoritma Genetika dengan *new fitness*, *roulette wheel selection*, dan *arithmetic crossover* berdasarkan *silhouette coefficient*.

E. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Dapat memberikan pengetahuan tentang penggunaan Algoritma Genetika dalam *clustering* data dengan *new fitness*, *roulette wheel selection*, *arithmetic crossover* dan *silhouette coefficient*.
2. Dapat menjadi sumber informasi data untuk penelitian selanjutnya.

