



Penerapan Algoritma Genetika Dalam Mengoptimasi *K-Means Clustering* Untuk Menentukan Prioritas Keluarga Sehat Di Puskesmas Doko

Inun Pujiyanto¹, Abdi Pandu Kusuma²,

^{1,2}Teknik Informatika, Universitas Islam Balitar. Blitar
inupiji89@gmail.com¹, pans.uibblitar@gmail.com²

Kata Kunci:

Prioritas Keluarga Sehat,
Algoritma genetika,
K-Means,
CARL

ABSTRAK

Pemerintah pusat mencanangkan Program Indonesia Sehat yang merupakan program utama yang ingin dicapai melalui Rencana Strategis (Renstra) Kementerian Kesehatan 2015-2019. Dalam hal ini Kementerian Kesehatan menetapkan strategi operasional dalam pembangunan kesehatan melalui Program Indonesia Sehat dengan Pendekatan Keluarga (PIS-PK). Didukung dengan adanya aplikasi berbasis *website* Keluarga Sehat yang digunakan untuk memasukkan hasil pendataan Keluarga Sehat. Aplikasi Keluarga Sehat dikembangkan untuk mempermudah proses pengolahan dan pengumpulan data pasien oleh institusi kesehatan daerah. Dalam penelitian ini, topik yang diangkat adalah optimalisasi metode *K-Means Clustering* untuk menentukan prioritas keluarga sehat di Doko. Terdapat 15 indikator keluarga sehat yang akan dibagi ke dalam 5 klaster. K-Means dan Algoritma Genetika dipilih karena prosesnya yang sederhana dengan komputasi yang ringan, metode yang digunakan adalah CARL. Hasil yang didapatkan Algoritma genetika dapat digunakan untuk mengoptimasi *K-Means clustering*. Hasil pengujian *Silhouette Coefficient* didapatkan nilai 0.9949 yang menunjukkan bahwa objek tersebut berada di dalam klaster yang sesuai dan memisahkan diri dengan baik dari klaster lainnya. Namun terdapat kekurangan yang menyebabkan proses untuk *preprocessing* belum cukup efektif karena harus melakukan perubahan dan pemilihan data yang sesuai dengan sistem yang digunakan.

ABSTRACT

The central government launched the Healthy Indonesia Program which is the main program to be achieved through the Ministry of Health's 2015-2019 Strategic Plan (Renstra). In this case the Ministry of Health establishes an operational strategy for health development through the Healthy Indonesia Program with a Family Approach (PIS-PK). Supported by the existence of a Healthy Families website-based application that is used to enter the results of Healthy Families data collection. The Healthy Family application was developed to facilitate the process of processing and collecting patient data by regional health institutions. In this study, the topic raised was the optimization of the *K-Means Clustering* method to determine the priority of healthy families at Doko. There are 15 healthy family indicators which will be divided into 5 clusters. K-Means and Genetic Algorithms were chosen because the process is simple with light computation, the method used is CARL. The results obtained by the genetic algorithm can be used to optimize *K-Means clustering*. The *Silhouette Coefficient* test results obtained a value of 0.9949 which indicates that the object is in the appropriate cluster and separates itself well from other clusters. However, there are deficiencies that cause the process for *preprocessing* to be not effective enough because it has to make changes and select data according to the system used

Keyword:

Healthy Families Priority,
Genetic Algorithm,
K-Means
CARL

PENDAHULUAN

Pemerintah pusat mencanangkan Program Indonesia Sehat yang merupakan program utama yang ingin dicapai melalui Rencana Strategis (Renstra) Kementerian Kesehatan 2015-2019. Sasarannya adalah meningkatnya derajat kesehatan melalui upaya kesehatan dan pemberdayaan masyarakat yang didukung dengan perlindungan finansial dan pemerataan pelayanan kesehatan. Dalam hal ini Kementerian Kesehatan menetapkan strategi operasional dalam pembangunan kesehatan melalui Program Indonesia Sehat dengan Pendekatan Keluarga (PIS-PK).

Pendekatan keluarga merupakan cara Puskesmas untuk meningkatkan jangkauan dan mendekatkan atau meningkatkan akses pelayanan kesehatan di ruang kerjanya dengan mengunjungi keluarga. Melalui program ini, puskesmas tidak hanya memberikan pelayanan kesehatan di dalam gedung, tetapi juga keluar gedung untuk mengunjungi keluarga di tempat kerja.

Aplikasi Keluarga Sehat dikembangkan untuk mempermudah proses pengolahan dan pengumpulan data pasien oleh institusi kesehatan daerah. Aplikasi Keluarga Sehat hanya dapat menampilkan indeks keluarga sehat. Parameter evaluasi pada aplikasi Keluarga Sehat didasarkan pada perhitungan 12 indikator keluarga. Indikator Keluarga Sehat tersebut antara lain:

- a. Seluruh keluarga sesuai definisi operasional mengikuti program Keluarga Berencana
- b. Pada ibu yang melakukan persalinan di fasilitas kesehatan
- c. Jika terdapat bayi, maka bayi tersebut mendapat imunisasi dasar lengkap
- d. Jika terdapat bayi, maka bayi tersebut diberi ASI eksklusif selama 6 bulan
- e. Bayi yang dipantau pertumbuhan dan perkembangannya setiap bulan
- f. Jika terdapat penderita TB Paru maka diobati sesuai standar
- g. Jika terdapat penderita darah tinggi atau hipertensi maka berobat teratur
- h. Jika terdapat penderita gangguan jiwa berat maka harus diobati sesuai standar dan tidak diterlantarkan
- i. Dalam satu keluarga tidak ada anggota keluarga yang merokok
- j. Satu keluarga harus mempunyai akses terhadap air bersih
- k. Satu keluarga mempunyai akses atau menggunakan jamban sehat
- l. Satu keluarga menjadi anggota JKN

Nilai indeks keluarga sehat dari 12 indikator meliputi: diatas 0,8 tergolong keluarga sehat, bila nilai indeks antara 0,5 sampai 0,8 tergolong keluarga pra sehat dan bila indeks kurang dari 0,5 tergolong keluarga tidak sehat. Menurut Teori klasik H. L. Bloom yang menyatakan bahwa ada 4 faktor yang mempengaruhi derajat kesehatan, yaitu: gaya hidup (*life style*), lingkungan (sosial, ekonomi, politik, budaya); pelayanan kesehatan; dan faktor genetik (keturunan) maka perlu adanya parameter tambahan dalam indikator keluarga sehat untuk dapat mengategorikan ke dalam kluster

yang sesuai. Indikator tambahan meliputi usia, tingkat pendidikan dan penghasilan. Sedangkan tingkatan klaster yang dimaksud terbagi menjadi 5 skala, diantaranya; klaster 1 (sangat prioritas), klaster 2 (prioritas), klaster 3 (sedang), klaster 4 (bukan prioritas) dan klaster 5 (sangat bukan prioritas)

Intervensi keluarga sehat adalah upaya untuk meningkatkan status indeks keluarga sehat dengan berbagai cara. Dengan terbatasnya kemampuan sumber daya manusia di puskesmas, anggaran dan waktu maka perlu dirangking ulang dari data keluarga sehat tersebut dengan menambahkan parameter indikator pendapatan dan pendidikan. Sehingga didapatkan data sasaran keluarga yang lebih objektif sesuai teori H.L. Bloom diatas, dengan demikian untuk mendapatkan hasil data yang efisien sesuai dengan target, penghitungan data tersebut akan dilakukan pengklasteran dengan menggunakan metode K-means clustering yang dioptimasi oleh penerapan algoritma genetika.

Algoritma Genetika (GA) adalah algoritma pencarian dan optimasi yang menggunakan prinsip evolusi alami. Secara umum, algoritma genetika terdiri dari dua proses. Proses pertama adalah seleksi individu untuk menghasilkan generasi selanjutnya dan proses kedua adalah manipulasi individu terpilih untuk membentuk generasi selanjutnya melalui teknik persilangan dan mutasi. Mekanisme seleksi menentukan individu mana yang akan dipilih untuk reproduksi dan berapa banyak keturunan yang akan dihasilkan oleh masing-masing dari individu tersebut. Sedangkan jumlah data yang besar perlu diintegrasikan menggunakan algoritma K-Mean untuk melihat karakteristik data keluarga sehat.

Optimalisasi atau optimasi yaitu suatu proses untuk mencapai hasil yang ideal atau optimasi (nilai efektif yang dapat dicapai). Optimasi dapat diartikan sebagai suatu bentuk mengoptimalkan sesuatu hal yang sudah ada, ataupun merancang dan membuat sesuatu secara optimal. Menurut Kamus Besar Bahasa Indonesia, “Optimalisasi adalah proses, cara dan perbuatan untuk mengoptimalkan (menjadikan paling baik, paling tinggi)”.

Optimalisasi adalah sebuah proses, cara dan perbuatan (aktivitas/kegiatan) untuk mencari solusi terbaik dalam beberapa masalah, dimana yang terbaik sesuai dengan kriteria tertentu. Dalam penelitian ini, topik yang diangkat adalah optimalisasi metode K-Means Clustering untuk menentukan prioritas keluarga sehat di Doko.

Metode k-mean merupakan metode akurasi yang baik untuk mengelompokkan data dalam jumlah besar [1]. Hal ini sesuai dengan data Keluarga Sehat yang ada di Puskesmas Doko dengan jumlah keluarga 13.387 KK per Januari 2023. Algoritma k-means memiliki kemampuan untuk mengelompokkan data dalam jumlah besar dengan waktu komputasi yang cepat, namun algoritma k-means memiliki kelemahan yaitu nilai awal dari pusat cluster bersifat acak yang sangat mempengaruhi hasil clustering. Selanjutnya iterasi yang didapatkan pada metode k-means sangat banyak untuk jumlah data yang digunakan. Oleh karena itu perlu dilakukan teknik optimasi untuk mempercepat proses clustering sekaligus memaksimalkan hasil [2].

Algoritma genetika adalah algoritma optimasi yang prinsip dasar komputasinya menggunakan seleksi alam melalui teori evolusi Darwin [3]. Algoritma genetika sangat baik dalam memecahkan

masalah kompleks dengan banyak variable [1]. Selain itu, algoritma genetika mampu mengelola optimasi yang berbeda tergantung pada fungsi tujuan (fitness) dan menjelajahi ruang pencarian yang cukup besar [4]

Penelitian ini merupakan salah satu bentuk penelitian terapan yang hasil penelitiannya merupakan jawaban atas permasalahan yang muncul dan digunakan sebagai solusi di instansi terkait. Penelitian terapan berlangsung dalam kaitannya dengan fakta praktis, penerapan dan pengembangan pengetahuan yang dihasilkan melalui penelitian dasar dalam kehidupan nyata.

Penelitian tentang clustering data dengan menggabungkan dua algoritma dan menunjukkan bahwa algoritma genetika dan k-means dapat menghasilkan tingkat clustering yang lebih baik daripada k-means tradisional [5].

Kombinasi kedua metode algoritma dapat diuji dengan silhouette coefficient. Silhouette coefficient merupakan metode pengujian sistem yang merupakan gabungan dari dua metode yaitu metode kohesi yang mengukur seberapa erat hubungan antar objek dalam suatu kelompok, sedangkan metode pemisahan yaitu mengukur jarak antara satu kelompok dengan kelompok lainnya. Kemudian hasil penerapan metode algoritma genetika untuk mengoptimalkan clustering k-means diuji dengan menggunakan metode silhouette coefficient yang dapat dievaluasi secara akurat sesuai dengan kluster yang ditentukan.

Seperti yang sudah dijelaskan di atas, peneliti bermaksud mengumpulkan data dari keluarga sehat dengan mengimplementasikan kombinasi algoritma genetika dan metode k-means di Puskesmas Doko. Hal ini dimaksudkan untuk memberikan informasi kepada Puskesmas tentang data keluarga prioritas yang perlu diintervensi terlebih dahulu agar program intervensi lebih terarah, objektif, efektif dan efisien.

Berdasarkan penjelasan sebelumnya yang disampaikan pada latar belakang, maka penulis menyusun skripsi dengan judul “Penerapan Metode Algoritma Genetika Dalam Mengoptimasi K-Means Clustering Untuk Menentukan Prioritas Keluarga Sehat di Puskemas Doko”

METODE

Variabel individu yang digunakan dalam penelitian ini meliputi kriteria keluarga mengikuti KB, Ibu melakukan persalinan di Faskes, Bayi mendapatkan imunisasi dasar lengkap, Bayi diberikan ASI eksklusif selama 6 bulan, Memantau pertumbuhan dan perkembangan balita tiap tahun, Penderita TB Paru berobat sesuai standar, Penderita hipertensi berobat teratur, Gangguan jiwa berat tidak ditelantarkan, Tidak ada anggota keluarga yang merokok, Keluarga mempunyai akses terhadap air bersih, Keluarga mempunyai atau menggunakan jamban sehat, Sekeluarga menjadi anggota JKN, Tingkat Pendidikan dan jenis pekerjaan keluarga, Penghasilan rata rata keluarga dan Usia produktif

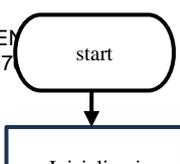
Metode penelitian yang digunakan dalam penelitian ini adalah metode R&D (Research and Development). Metode penelitian dan pengembangan atau dalam bahasa Inggrisnya Research and

Development adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan menguji keefektifan produk tersebut [6]

Produk yang dihasilkan dari penelitian ini adalah metode clustering dari data keluarga sehat dengan menggunakan algoritma genetika dan K-Mens clustering. Metode yang digunakan pada penelitian ini adalah Algoritma Genetika dan K-Means Clustering. Pada tahap awal pre-processing dilakukan penentuan bobot untuk setiap indikator dengan metode CARL berdasarkan bobot yang telah diperoleh. Sehingga data yang semua bernilai N, Y dan T, berubah menjadi nilai bobot dalam angka. Selanjutnya dari 15 indikator diklastering menjadi 5 klaster berdasarkan rangking tertinggi hingga terendah.

Cluster ditentukan kemudian inialisasi parameter. Inialisasi parameter adalah banyak individu (popSize), banyak generasi AG, crossover rate, dan mutation rate. Gambar 1.1 adalah diagram alir dari perancangan algoritme AG - K-Means. Gambar 1.1 merupakan tahap pertama dari metode yang diimplementasikan adalah menentukan inialisasi populasi kemudian menentukan jumlah cluster. Inialisasi populasi yaitu menghasilkan beberapa individu awal. Pada penelitian ini, banyak cluster ditentukan setelah melakukan inialisasi populasi adalah 5 cluster. Panjang kromosom sebanyak 15. Reproduksi crossover dilakukan secara acak memilih 50 individu awal dan membangkitkan secara acak nilai alpha. Banyaknya offspring yang dihasilkan sebanyak crossover rate ukuran populasi (popSize) Reproduksi mutation dilakukan dengan memilih 4 individu secara acak untuk selanjutnya diproses dan membangkitkan secara acak nilai r. Banyaknya offspring yang dihasilkan sebanyak mutation rate ukuran populasi (popSize).

Pada tahap penerapan K-means clustering pertama inisiasi jumlah klaster yang diinginkan sebanyak 5 klaster bersadarkan hasil optimasi algoritma genetika, hitung jarak antara data dengan pusat klaster, kemudian kelompokkan data dengan berdasarkan jarak terdekat ke pusat klaster, ulangi langkah-langkah tersebut sampai tidak ada pengelompokan data yang berpindah klaster dari suatu klalster/kelompok. Design tahap penelitian dapat dilihat pada digram berikut:



Gambar 1 Diagram proses penelitian

Data sampel yang digunakan dalam penelitian ini adalah sebanyak 50 data Keluarga di Puskesmas Doko Kabupaten Bitar.

Teknik pengumpulan data

Teknik pengumpulan data dalam penelitian ini menggunakan metode survey ke Puskesmas, observasi ke Puskesmas, wawancara Kepala Puskesmas dan Programmer, serta studi literasi yang didapatkan dari data aplikasi keluarga sehat.

Teknik analisis data

Hasil data setelah dilakukan clustering maka dilakukan pengujian dengan metode silhouette coefficient untuk melihat seberapa nilai seberapa bagus setiap objek terletak dalam suatu klaster/kelompok

HASIL DAN PEMBAHASAN

Hasil

A. Dataset

Data yang didapatkan dari Puskesmas Doko merupakan data mentah yang belum digolongkan secara detail ke dalam 15 indikator kriteria keluarga sehat. Data tersebut masih perlu melewati berbagai penjarangan lebih dalam. Data mentah yang didapatkan dari Puskesmas Doko berupa identitas lengkap 37463 masyarakat yang dikelompokkan menjadi masing-masing KK. Data tersebut dapat dilihat pada table 1.1.

Tabel 1. Data Sampel Keluarga Sehat Puskesmas Doko

NAMA KK	NAMA ART	USIA	MENGGUNAKAN KB	PERSALINAN DI FASKES	IMUNISASI LENGKAP	ASI EKSKLUSIF	PEMANTAUAN PERTUMBUHAN BALITA	MINUM OBAT TB TERATUR	MINUM OBAT HIPERTENSI TERATUR	ART MINUM OBAT GANGGUAN JIWA BERAT TERATUR	MEROKOK	PERILAKU PENGGUNAAN AIR BERSIH	PERILAKU BAB DI JAMBAN	KEPESEKERTAAAN JKN	PENGHASILAN INDIVIDU	PENDIDIKAN
ARI SETIAWAN	ARI SETIAWAN	24	N	N	N	N	N	N	N	N	T	Y	Y	Y	2500000	SLTP/Sederajat Belum Tamat
	MURYATI	54	N	N	N	N	N	N	N	N	Y	Y	Y	N	600000	SD/Sederajat
AGUS SUSANTO	IKA NUR TISYA LINDRIA	21	N	N	N	N	N	N	N	N	Y	Y	Y	N	0	SLTP/Sederajat Belum Tamat
	GUSMIRANDA AGUS SUSANTO	9	N	N	N	N	N	N	N	N	Y	Y	Y	T	0	SD/Sederajat
MESIDI YUSUF	MESIDI YUSUF	29	N	N	N	N	N	N	N	N	T	Y	Y	Y	2000000	Tamat SD/Sederajat
	SRIATUN	52	N	N	N	N	N	N	N	N	T	Y	Y	Y	3500000	SLTA/Sederajat
SISWANTO	DINDA O LUTUT	56	N	N	N	N	N	N	N	N	T	Y	Y	Y	0	Tamat SD/Sederajat Belum Tamat
	PUJIASTUTIK	11	N	N	N	N	N	N	N	N	T	N	N	T	0	SD/Sederajat
BEJO HADI PRASONO	SISWANTO	34	Y	N	N	N	N	N	N	N	T	Y	Y	T	1000000	Tamat SD/Sederajat
	DENI ADITYA	36	Y	N	N	N	N	N	N	N	T	Y	Y	T	2000000	Tamat SD/Sederajat
	KARNI RAMADHANU IGOR	14	N	N	N	N	N	N	N	N	T	N	N	T	0	Tamat SD/Sederajat Tidak/Belum Sekolah
	KHAFKANA ENI RATNAWATI	86	N	N	N	N	N	N	N	N	Y	Y	T	T	1800000	Belum Tamat SD/Sederajat
	ZIEGO RAMADEFANO	10	N	N	N	N	N	N	N	N	T	N	N	T	0	Belum Tamat SD/Sederajat
BEJO HADI PRASONO	44	N	N	N	N	N	N	N	N	T	Y	T	T	2000000	SLTP/Sederajat	
KARTINI	16	N	N	N	N	N	N	N	N	T	Y	T	T	0	SLTP/Sederajat	
		45	N	N	N	N	N	N	N	N	T	Y	T	T	2500000	Tamat SD/Sederajat Tidak/Belum Sekolah
		75	N	N	N	N	N	N	T	N	T	Y	T	T	0	Tamat SD/Sederajat Tidak/Belum Sekolah

Keterangan : N=Nihil, Y=ya, T=Tidak

Seluruh data tersebut telah diberi keterangan menggunakan Y (Ya) T (Tidak) N (Netral) sesuai dengan kelengkapan data yang ada dari masing-masing KK. Selanjutnya seluruh data tersebut dihitung menggunakan CARL, adapun hasil penghitungannya sebagai berikut:

Tabel 2. Hasil pembobotan indikator dengan CARL

Variabel	Keterangan	C	A	R	L	Hasil bobot	Skala
X1	Keluarga mengikuti KB	3	2	2	2	24	0,046
X2	Ibu melakukan persalinan di Faskes	4	1	2	1	8	0,015
X3	Bayi mendapatkan imunisasi dasar lengkap	3	2	2	2	24	0,046
X4	Bayi diberi ASI eksklusif selama 6 bulan	3	3	2	1	18	0,035
X5	Memantau pertumbuhan dan perkembangan balita tiap bulan	3	2	1	3	18	0,035
X6	Penderita TB dan paru berobat sesuai dengan standar	3	3	2	4	72	0,138
X7	Penderita hipertensi berobat dengan teratur	4	4	2	4	128	0,246
X8	Gangguan jiwa berat tidak ditelantarkan	3	3	3	2	54	0,104
X9	Tidak ada anggota keluarga yang merokok	3	1	2	2	12	0,023

X10	Keluarga mempunyai akses air bersih	2	2	2	4	32	0,062
X11	Keluarga mempunyai akses atau menggunakan jamban sehat	2	3	2	4	48	0,092
X12	Sekeluarga menjadi anggota JKN	2	3	1	1	6	0,012
X13	Jenis Pekerjaan Keluarga/penghasilan keluarga	2	2	2	3	24	0,046
X14	Pendidikan terakhir Keluarga	2	2	2	3	24	0,046
X15	Usia individu	2	2	2	2	16	0,031

Setelah dilakukan penghitungan menggunakan CARL, dilanjutkan dengan melakukan pengelompokan data yang telah dihitung ke dalam 5 kriteria, kriteria tersebut akan diurutkan sesuai dengan nilai tertingginya. Berikut ini merupakan hasil dan definisi operasional untuk kriteria yang dibuat:

Tabel 3 Tabel Definisi operasional pengelompokan Indikator prioritas

Kriteria	Pengertian
C1	-Penderita hipertensi berobat dengan teratur -Gangguan jiwa berat tidak ditelantarkan
C2	-Penderita TB dan paru berobat sesuai dengan standar -Jenis pekerjaan keluarga/penghasilan keluarga -Keluarga mempunyai akses air bersih -Bayi diberi ASI eksklusif selama 6 bulan
C3	-Memantau pertumbuhan dan perkembangan balita tiap bulan -Pendidikan terakhir keluarga -Keluarga mengikuti KB
C4	-Bayi mendapatkan imunisasi dasar lengkap -Keluarga mempunyai akses atau menggunakan jamban sehat -Usia individu
C5	-Tidak ada anggota keluarga yang merokok -Sekeluarga menjadi anggota JKN -Ibu melakukan persalina di Faskes

Dataset yang sudah siap diolah

Tabel 4 Tabel Dataset yang siap diproses

NAMA KK	c1	c2	c3	c4	c5
ARI SETIAWAN	0	0.03	0.267	0.111	0.01
AGUS SUSANTO	0	0	0.2	0.031	0.017
MESIDI YUSUF	0	0	0.1	0.031	0.02
JIYEM	0.321	0.027	0.1	0.031	0.017
SISWANTO	0	0	0.4	0.031	0.068
BEJO HADI PRASONO	0.321	0.135	0.54	0.031	0.092
YULIANTO	0	0	0.26	0.031	0.041
JIANTO	0	0	0.3	0.031	0.031
GIYAH	0	0	0.34	0.031	0.048

B. Proses Implementasi Perhitungan

Dalam mengimplementasikan perhitungan terdapat beberapa tahap yang digunakan, berikut ini merupakan proses yang digunakan.

1) Tahap Inisialisasi

Dalam tahap ini jumlah populasi kromosom yang akan digunakan dalam algoritma genetika ditentukan berdasarkan jumlah data yang ada. Tahap ini menggunakan teknik heuristik, teknik heuristik adalah teknik yang digunakan untuk menemukan solusi atau membuat

keputusan dengan cepat dalam situasi yang kompleks atau tidak terstruktur. Heuristik didasarkan pada pengalaman, pengetahuan domain, aturan praktis, atau intuisi untuk membantu menghasilkan solusi yang memadai dalam waktu yang terbatas.

2. Tahap Evaluasi

Tahap ini merupakan jarak antara setiap individu dengan centroid dihitung menggunakan metrik jarak, seperti Euclidean distance, dengan memanggil fungsi Euclidean Distance.

3. Tahap Seleksi

Tahap seleksi dalam konteks penerapan algoritma k-means adalah proses dimana kita memilih titik titik pusat (centeroid) yang tepat untuk mewakili kelompok- kelompok dalam kumpulan data.

4. Tahap Reproduksi

Tahap Reproduksi merupakan tahap yang penting dalam algoritma genetika. Tahap ini mencakup proses reproduksi atau pembiakan populasi saat mencari solusi terbaik untuk masalah yang sedang dipecahkan.

5. Tahap Iterasi

Tahap iterasi digunakan untuk melakukan perhitungan berulang dengan tujuan mendekati atau mencapai solusi yang diinginkan.

Di setiap iterasi, dilakukan perhitungan total fitness, pembuatan kromosom-kromosom baru sebagai keturunan melalui operasi crossover dan mutasi, penggantian populasi kromosom lama dengan kromosom-kromosom baru, serta penghitungan kluster dengan menggunakan algoritma K-Means pada populasi kromosom yang baru. Hasil kluster dan total jarak rata-rata ditampilkan dalam setiap iterasi menggunakan elemen HTML yang sesuai.

Analisa menggunakan metode Shilouette Coefficient Untuk mendapatkan nilai SC adalah dengan menghitung jarak rata-rata dari suatu data-*i* di satu kluster (a_i) dengan menggunakan Persamaan dengan memisalkan *i* terhadap seluruh data lain pada satu kluster.

Data kluster yang digunakan untuk pengujian adalah kluster 3, 4 dan 5, dan hasil penghitungan jarak terhadap centeroid dapat dilihat pada table berikut:

Tabel. 5 Hasil penghitungan jarak terhadap centeroid

ID	C1	C2	C3	C4	C5	Cluster	d c1	d c2	d c3	d c4	d c5
YULIAN		0.12	0.67	0.03	0.76			0.1838	0.6846	0.1414	0.7730
TO	0	15	06	1	06	3	0.138	65	52	39	18
SARBA	0.00	0.27		0.17	0.30		0.1381	0.3109	0.1965	0.2192	0.3329
NU	64	87	0.14	04	3	3	48	95	81	72	46
MULYO	0.04				0.56		0.1461		0.2429	0.4420	0.5862
NO	82	0	0.2	0.42	98	3	75	0.138	9	9	73
LASMIN		0.35	0.01	0.10	0.61			0.3730	0.1005	0.1472	0.6194
I	0	94	06	81	13	4	0.1	53	6	6	25
ARI											
SETIAW			0.14	0.18		5		0.0583	0.1513	0.1954	0.0509
AN	0	0.03	29	89	0.01		0.05	1	95	05	9

Kemudian lakukan penghitungan dengan persamaan berikut

$$a_i = \frac{1}{m-1, r \neq i} \sum m d(x_i, x_r), \quad i = 1, 2, \dots, m \quad \text{dan}$$

$$b_i = \min \frac{1}{m, r \neq i} \sum m d(x_i, x_r), \quad i = 1, 2, \dots, m$$

Kemudian hitung nilai SC pada data ke- i (SC_i), nilai SC pada suatu kluster (SCK), dan nilai SC pada seluruh pengklasteran (SC) menggunakan persamaan sebagai berikut

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}$$

Maka hasil penghitungan dengan rumus persamaan diatas dapat dilihat pada table berikut.

Tabel 6 Hasil penghitungan silhouette coefficient

ID	a(i) jarak	b(i) jarak	s(i)
YULIANTO	0.384194698	7.246376812	0.946981132
SARBANU	0.239588321	7.238596575	0.966901275
MULYONO	0.311105716	7.246376812	0.957067411
LASMINI	0.268059725	10	0.973194027
ARI SETIAWAN	0.101219968	20	0.994939002

Maka nilai $s(i)$ rata rata pada data tersebut adalah 0.9949 sehingga dapat dikatakan bahwa data tersebut sudah struktur kluster dinyatakan kuat atau berada pada kluster yang tepat.

Pembahasan

Dalam hasil klustering dengan metode K-Means yang dioptimasi menggunakan algoritma genetika, terdapat beberapa komponen penting yang dapat dijelaskan secara deskriptif sebagai berikut:

Pertama, konfigurasi klustering meliputi parameter-parameter yang digunakan dalam proses klustering. Jumlah populasi sebanyak 50 mengindikasikan bahwa ada 50 individu yang terbentuk dalam populasi pada setiap iterasi. Jumlah kluster sebanyak 5 menunjukkan bahwa klustering akan menghasilkan 5 kelompok yang berbeda. Terdapat juga 5 fitur yang digunakan dalam analisis ini. Selanjutnya, terdapat 15 iterasi yang menunjukkan berapa kali klustering dilakukan dan dievaluasi. Probabilitas crossover sebesar 0.7 dan probabilitas mutasi sebesar 0.6 menunjukkan tingkat kemungkinan terjadinya operasi crossover dan mutasi dalam algoritma genetika.

Kemudian, hasil klustering untuk setiap iterasi ditampilkan dengan prioritas 1 hingga prioritas 5. Prioritas tersebut menandakan tingkat kepentingan masing-masing kelompok kluster, di mana prioritas 1 adalah yang tertinggi. Setiap kelompok kluster memiliki pusat kluster (c_1, c_2, c_3, c_4, c_5) dengan nilai koordinat yang diberikan. Nama-nama individu (NAMA KK) kemudian diklasifikasikan ke dalam kelompok kluster yang sesuai berdasarkan prioritas terendah yang terpenuhi. Hal ini membantu dalam memahami bagaimana data dikelompokkan dan mengidentifikasi kluster mana yang memiliki tingkat kepentingan yang lebih tinggi.

Selanjutnya, total jarak rata-rata digunakan untuk mengevaluasi kualitas klustering yang dihasilkan. Semakin rendah nilai total jarak rata-rata, semakin baik kualitas klusteringnya. Dalam hasil yang diberikan, terdapat total jarak rata-rata pada setiap iterasi. Pada iterasi ke-14, diperoleh

total jarak rata-rata sebesar 33.737269339615, sedangkan pada iterasi ke-15, terjadi penurunan nilai total jarak rata-rata menjadi 31.68072163158. Hal ini menunjukkan adanya perbaikan dalam pembentukan kluster dari iterasi sebelumnya.

Selain itu, koefisien siluet juga memberikan informasi tentang kualitas klustering secara individu. Koefisien siluet berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa data tersebut cocok dengan kluster tempatnya berada. Dalam hasil yang diberikan, setiap data (NAMA KK) disertai dengan koordinatnya dan diikuti oleh nilai koefisien siluetnya. Informasi ini membantu dalam memahami sejauh mana setiap data cocok dengan kluster tempatnya ditempatkan.

Secara keseluruhan, hasil klustering ini memberikan pemahaman tentang pengelompokan data berdasarkan prioritas dan kualitas klustering yang dihasilkan. Melalui konfigurasi klustering, hasil iterasi, total jarak rata-rata, dan koefisien siluet, kita dapat menginterpretasikan bagaimana data dikelompokkan dan sejauh mana kualitas klustering yang dicapai. Penjelasan ini membantu dalam memahami proses klustering yang dilakukan dan memberikan wawasan yang lebih baik tentang pola-pola yang muncul dalam data tersebut.

Dampak dari penggunaan algoritma genetika dalam optimasi klustering dapat dilihat dari beberapa aspek berikut:

1) Perubahan Kualitas Klustering

Hasil klustering yang diperoleh dari iterasi ke-14 menunjukkan bahwa kualitas klustering telah meningkat dibandingkan dengan hasil awal (sebelum proses optimasi). Hal ini dapat dilihat dari penurunan total jarak rata-rata dari 33.737 pada awal iterasi menjadi 31.681 pada iterasi ke-15. Penurunan total jarak rata-rata ini menunjukkan bahwa kluster yang dihasilkan lebih mendekati titik data dan memiliki pemusatan yang lebih baik.

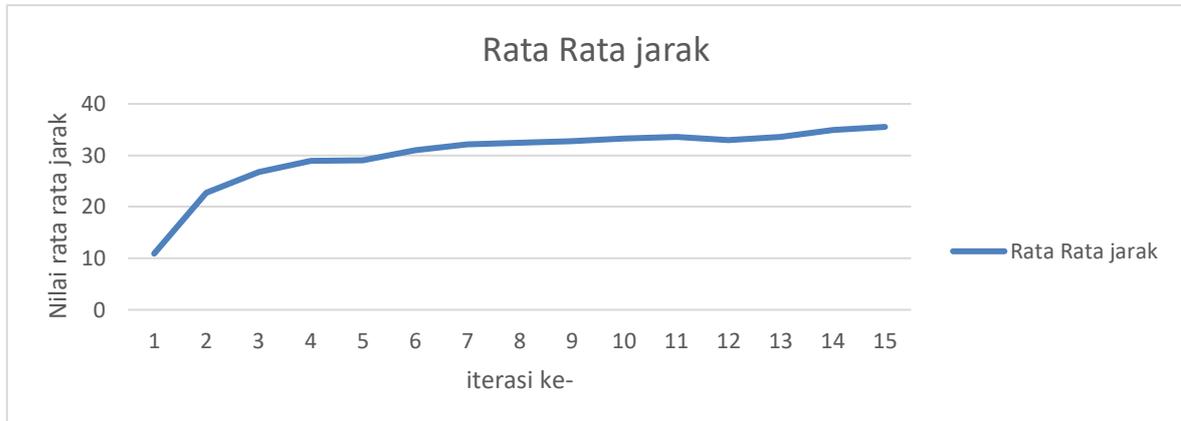
2) Silhouette Coefficient

Pada hasil perhitungan, terdapat perhitungan koefisien siluet untuk setiap data. Koefisien siluet digunakan untuk mengukur sejauh mana titik data cocok dengan kluster tempatnya ditempatkan. Semakin tinggi nilai koefisien siluet, semakin baik penempatan titik data dalam kluster yang sesuai. Dalam hasil tersebut, sebagian besar data memiliki koefisien siluet bernilai 1, yang menunjukkan kualitas klustering yang sangat baik.

Dari hasil perhitungan tersebut, dapat disimpulkan bahwa algoritma genetika telah memberikan dampak yang signifikan dalam meningkatkan kualitas klustering. Melalui proses optimasi menggunakan algoritma genetika, total jarak rata-rata antara data dan pusat kluster berhasil diperkecil, menunjukkan adanya peningkatan pemusatan data dalam kluster yang lebih baik. Selain itu, koefisien siluet yang tinggi pada sebagian besar data menunjukkan bahwa penempatan titik data dalam kluster telah sesuai dan memperkuat validitas hasil klustering.

Hal ini menunjukkan bahwa algoritma genetika telah berhasil mengoptimalkan konfigurasi kluster dengan mempertimbangkan jarak antara data dan pusat kluster. Penggunaan algoritma genetika dalam optimasi klustering memberikan dampak yang signifikan dalam meningkatkan kualitas klustering dan memberikan solusi yang lebih baik dalam pemusatan data ke dalam kluster

yang sesuai. Hasil penghitungan nilai rata – rata jarak terhadap centeroid pada setiap kluster terhadap jumlah iterasi dapat dilihat pada grafik berikut:



Gambar 2 Jarak rata rata terhadap centeroid pada iterasi

Dari gambar 1.2 diatas dapat disimpulkan data mulai stabil menempati klasternya setelah dilakukan iterasi ke 10 hingga seterusnya.

SIMPULAN

Adapun kesimpulan yang dapat diambil pada pembahasan bab IV dari penelitian ini adalah

1. Algoritma genetika dapat digunakan untuk mengoptimasi K-Means clustering. algoritma genetika telah berhasil mengoptimalkan konfigurasi kluster dengan mempertimbangkan jarak antara data dan pusat kluster. Penggunaan algoritma genetika dalam optimasi klustering memberikan dampak yang signifikan dalam meningkatkan kualitas klustering dan memberikan solusi yang lebih baik dalam pemusatan data ke dalam kluster yang sesuai.
2. Identifikasi nilai Silhouette Coefficient yang didapatkan adalah nilai 0.9949 sebagaimana terlihat di tabel 1.6. Hal tersebut menunjukkan bahwa objek tersebut memiliki kecocokan dengan kelompoknya dengan baik dan memiliki jarak yang signifikan dengan kelompok lain. Dengan demikian berarti bahwa objek tersebut berada di dalam kluster yang sesuai dan memisahkan diri dengan baik dari kluster lainnya.

DAFTAR RUJUKAN

- [1] S. Santoso, “Menguasai Statistik Dengan SPSS 24” Jakarta: PT Alex Media Komputindo, 2017.
- [2] M.R. Zeebaree, dan Siron, R. B. 2017, “The Impact of Entrepreneurial Orientation on Competitive Advantage Moderated by Financing Support in SMEs”, *International Review of Management and Marketing*, Volume 7, No.1
- [3] Mursalim, “Organizational Culture and Organizational Climate as a Determinant of Motivation and Teacher Performance.” *Advances in Social Sciences Research Journal* 6, no. 2, 2021
- [4] Prasetio, “Pengaruh Kepuasan Kerja, Lingkungan Kerja, dan Kompensasi Terhadap Turnover Intention Karyawan dengan Lama Kerja Sebagai Variabel Pemoderasi TURNOVER INTENTION (Studi Kasus Pada CV. Sukses Sejati Computama)”, 2018

- [5] W. Khotimah N., “Platform E-Learning untuk Pembelajaran Pemrograman Web Menggunakan Konsep Progressive Web Apps” JURNAL TEKNIK ITS Vol. 6, No. 2, 2017
- [6] Sugiyono, “Metode Penelitian Kombinasi (Mix Methods)”, Bandung, Alfabeta, 2015