

**EVALUASI HASIL *CLUSTERING* MENGGUNAKAN
ALGORITMA GENETIKA DENGAN *NEW FITNESS*,
ROULETTE WHEEL SELECTION DAN *ARITHMETIC
CROSSOVER***



SKRIPSI

**SILVIA NOVERA. M
1803040153**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
AGUSTUS 2021**

**ALGORITMA GENETIKA DENGAN *NEW FITNESS*,
ROULETTE WHEEL SELECTION DAN *ARITHMETIC
CROSSOVER* UNTUK *CLUSTERING DATA***



SKRIPSI

**diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana
Teknik/Komputer**

**SILVIA NOVERA.M
1803040153**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
AGUSTUS 2022**

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertandatangan dibawah ini:

Nama : Silvia Novera. M

NIM : 1803040153

Program Studi : Teknik Informatika

Fakultas : Teknik dan Sains

Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

Menyatakan dengan sebenar-benarnya bahwa skripsi ini adalah hasil karya saya dengan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta bukan hasil penjiplakan karya orang lain.

Demikian pernyataan ini saya buat dan apabila kelak kemudian hari terbukti ada unsur penjiplakan, saya bersedia mempertanggungjawabkan sesuai dengan ketentuan yang berlaku.

Purwokerto, 16 Juni 2022
Yang membuat pernyataan



Silvia Novera. M

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Purwokerto dan demi pengembangan ilmu pengetahuan, saya yang bertanda tangan di bawah ini:

Nama : Silvia Novera.M
NIM : 1803040153
Program Studi : Teknik Informatika
Fakultas : Teknik dan Sains
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto
Jenis Karya : Skripsi

menyetujui untuk memberikan Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right) kepada Universitas Muhammadiyah Purwokerto atas karya ilmiah saya yang berjudul:

Evaluasi Hasil *Clustering* Menggunakan Algoritma Genetika dengan *New Fitness, Roulette Wheel Selection* dan *Arithmetic Crossover*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Muhammadiyah Purwokerto berhak menyimpan, mengalihmedia/ mengalihformatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya dengan tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Purwokerto

Pada tanggal : 11 Agustus 2022

Yang menyatakan,

Silvia Novera. M



HALAMAN PERSEMBAHAN

“Segala perjuangan dan pengorbanan hingga titik ini dipersembahkan kepada dua orang yang paling berharga dan berarti dalam hidup yaitu kedua orang tua saya, Bapak Mukhlis dan Ibu Nurhailis, kakak saya Skeni Rahmawati dan adik saya Amel Amalia.M.”



HALAMAN MOTTO

“Happy parents happy life”

(Silvia Novera. M)



ABSTRAK

Clustering merupakan teknik pengenalan pola yang populer dan telah banyak digunakan dalam berbagai bidang, seperti *web mining*, segmentasi citra, *machine learning*, pengenalan biometrik, teknik elektro, teknik mesin, penginderaan jauh, dan genetika. AG salah satu metode untuk *clustering* yang bersifat pencarian global. AG dengan memperhatikan parameter yang digunakan memperoleh hasil yang optimal. Metode *clustering* AG diuji menggunakan 5 dataset. *Silhouette coefficient* digunakan sebagai evaluasi kualitas hasil *clustering*. Evaluasi dilakukan untuk mengecek seberapa optimal suatu *cluster*. Penelitian dilakukan untuk pengembangan AG dengan *new fitness*, *roulette wheel selection* dan *arithmetic crossover*. Tujuan penelitian dilakukan untuk mengecek seberapa baik hasil *clustering* menggunakan Algoritma Genetika dengan *new fitness*, *roulette wheel selection* dan *arithmetic crossover* berdasarkan *silhouette coefficient*. Pengujian dilakukan dengan dua skema yaitu kombinasi generasi 50 Cr=0.7, Cr=0.8, Cr=0.9 dan generasi 100 Cr=0.7, Cr=0.8, Cr=0.9. Hasil penelitian menunjukkan nilai evaluasi hasil *clustering* menggunakan AG berdasarkan *silhouette coefficient* paling tinggi yaitu kombinasi generasi 100 Cr=0.9 dengan nilai evaluasi setiap dataset yaitu iris = 0.88431, wine = 0.91681, glass = 0.80641, Ecoli = 0.677 dan WDBC = 0.8668.

Kata Kunci: *Clustering*, Algoritma Genetika, *new fitness*, *silhouette coefficient*.

ABSTRACT

Clustering is a popular pattern recognition technique and has been widely used in various fields, such as web mining, image segmentation, machine learning, biometric recognition, electrical engineering, mechanical engineering, remote sensing, and genetics. AG is one of the methods for global search clustering. AG by paying attention to the parameters used to obtain optimal results. The AG clustering method was tested using 5 datasets. Silhouette coefficient is used as an evaluation of the quality of clustering results. Evaluation is done to check how optimal a cluster is. The research was conducted for the development of AG with new fitness, roulette wheel selection and arithmetic crossover. The purpose of this research is to check how well the results of clustering using Genetic Algorithm with new fitness, roulette wheel selection and arithmetic crossover based on silhouette coefficient. The test was carried out using two schemes, namely a combination of generation 50 $Cr=0.7$, $Cr=0.8$, $Cr=0.9$ and generation 100 $Cr=0.7$, $Cr=0.8$, $Cr=0.9$. The results show the evaluation value of clustering results using AG based on the highest silhouette coefficient, namely the combination of generation 100 $Cr = 0.9$ with the evaluation value of each dataset, namely iris = 0.88431, wine = 0.91681, glass = 0.80641, Ecoli = 0.677 and WDBC = 0.8668.

Keywords: Clustering, Genetic Algorithm, new fitness, silhouette coefficient.



KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan karuniaNya kepada penulis, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Shalawat dan salam senantiasa tercurah kepada Rasulullah SAW yang mengantarkan manusia dari zaman jahiliah ke zaman yang penuh dengan ilmu pengetahuan ini. Penulis menyadari penulisan skripsi ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun materil. Oleh sebab itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

1. Apak, terimakasih atas segala doa, nasehat dan support yang tidak pernah dapat ku balas hanya dengan selembar kertas.
2. Amak, terimakasih selalu mendoakanku, selalu sabar mendengarkan segala keluh kesahku dan memberiku motivasi untuk selalu kuat.
3. Kakak dan Adikku Skeni Rahmawati dan Amel Amalia. M terimakasih selalu memberiku semangat dan motivasi untuk tetap sabar dalam berjuang.
4. Feri Wibowo, S.Kom., M.Cs. selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Purwokerto.
5. Maulida Ayu Fitriani, S.Kom., M.Cs. selaku dosen Pembimbing Skripsi terimakasih telah memberikan bimbingan dan arahan dengan penuh kesabaran selama proses penelitian dan penulisan skripsi ini.
6. Serta semua pihak yang tidak dapat disebut satu per satu yang telah membantu penyusunan dalam menyelesaikan laporan ini.

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari sempurna dikarenakan terbatasnya pengetahuan dan pengalaman yang dimiliki penulis. Oleh karena itu penulis mengharapkan segala bentuk saran dan masukan serta kritikan yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak khususnya dalam bidang data mining.

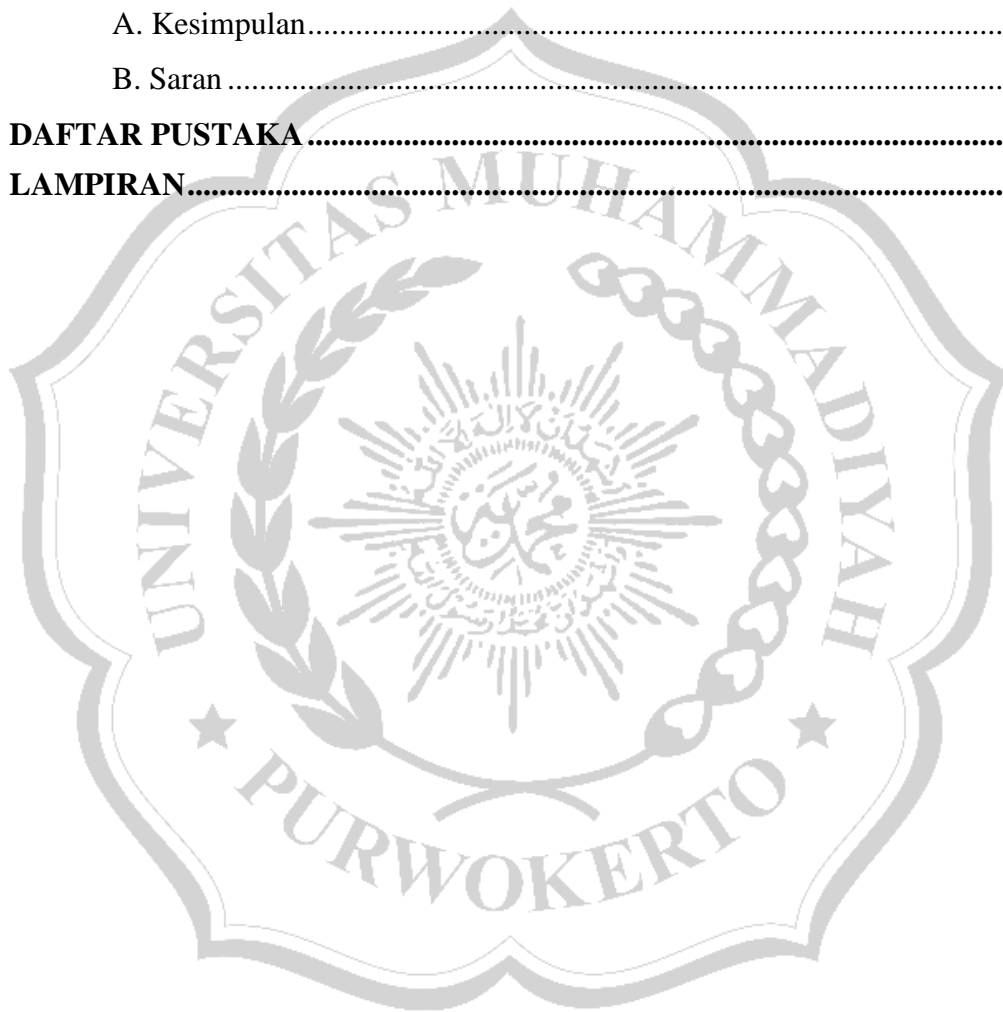
Purwokerto, 27 Juli 2022

Silvia Novera.M

DAFTAR ISI

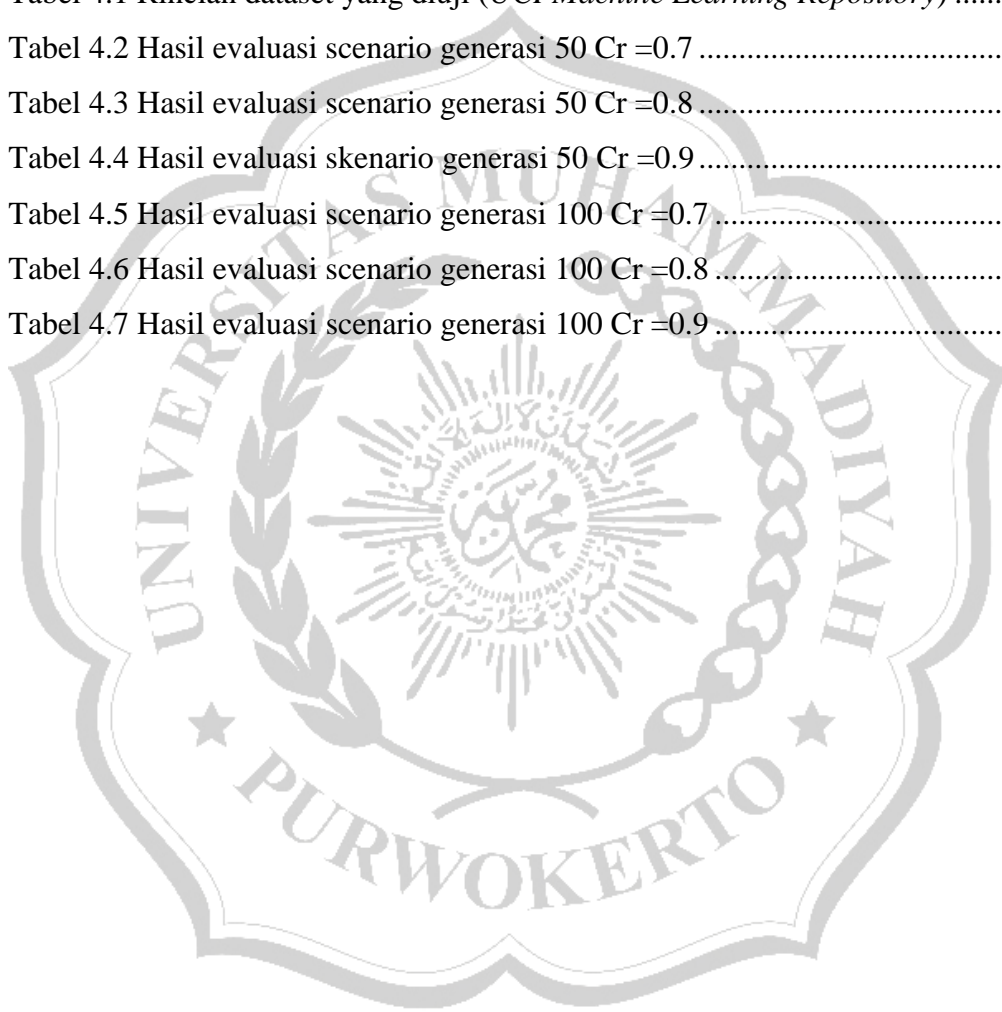
SKRIPSI	i
SKRIPSI	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN ORISINALITAS	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	vi
HALAMAN PERSEMBAHAN	vii
HALAMAN MOTTO	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR	xi
DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Perumusan Masalah.....	2
C. Batasan Masalah.....	2
D. Tujuan Penelitian.....	3
E. Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
A. Penelitian Terdahulu	4
B. Kajian Teori.....	7
BAB III METODE PENELITIAN	16
A. Jenis Penelitian.....	16
B. Waktu dan Tempat Penelitian	16
C. Metode Pengumpulan Data	16
D. Data Uji	16
E. Arsitektur Penelitian	17

BAB IV HASIL DAN PEMBAHASAN	22
A. Analisis Kebutuhan	22
B. Implementasi Metode <i>Clustering</i>	23
C. Arsitektur Evaluasi Hasil <i>Clustering</i>	33
D. Pengujian Hasil <i>Clustering</i>	35
BAB V PENUTUP.....	50
A. Kesimpulan.....	50
B. Saran	50
DAFTAR PUSTAKA.....	51
LAMPIRAN.....	53



DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	4
Tabel 2.1 (Lanjutan).....	5
Tabel 2.1 (Lanjutan).....	6
Tabel 3.1 Dataset.....	16
Tabel 4.1 Rincian dataset yang diuji (<i>UCI Machine Learning Repository</i>)	22
Tabel 4.2 Hasil evaluasi scenario generasi 50 Cr =0.7	36
Tabel 4.3 Hasil evaluasi scenario generasi 50 Cr =0.8	38
Tabel 4.4 Hasil evaluasi skenario generasi 50 Cr =0.9	40
Tabel 4.5 Hasil evaluasi scenario generasi 100 Cr =0.7	43
Tabel 4.6 Hasil evaluasi scenario generasi 100 Cr =0.8	45
Tabel 4.7 Hasil evaluasi scenario generasi 100 Cr =0.9	47



DAFTAR GAMBAR

Gambar 3.1	Arsitektur penelitian.....	17
Gambar 4.1	<i>Source code</i> inialisasi kromosom	24
Gambar 4.2	<i>Source code</i> pencarian jarak terdekat.....	25
Gambar 4.3	<i>Source code</i> evaluasi hasil <i>clustering</i>	26
Gambar 4.4	<i>Source code</i> perhitungan jarak total antar <i>cluster</i> (BC).....	27
Gambar 4.5	<i>Source code</i> perhitungan jarak total data dalam satu <i>cluster</i> (WC).....	28
Gambar 4.6	<i>Source code</i> perhitungan lebar <i>silhouette</i> (SW).....	29
Gambar 4.7	<i>Source code</i> proses seleksi	30
Gambar 4.8	<i>Source code</i> proses <i>crossover</i>	31
Gambar 4.9	<i>Source code</i> proses mutasi	32
Gambar 4.10	<i>Source code</i> perbandingan nilai evaluasi	33
Gambar 4.11	Arsitektur pengujian.....	34
Gambar 4.12	Grafik evaluasi maksimal setiap dataset skenario 1a.....	37
Gambar 4.13	Grafik evaluasi maksimal setiap dataset skenario 1b.....	39
Gambar 4.14	Grafik evaluasi maksimal setiap dataset skenario 1c.....	41
Gambar 4.15	Grafik nilai evaluasi tertinggi <i>silhouette coefficient</i> skema 1	42
Gambar 4.16	Grafik evaluasi maksimal setiap dataset skenario 2a.....	44
Gambar 4.17	Grafik evaluasi maksimal setiap dataset skenario 2b.....	46
Gambar 4.18	Grafik evaluasi maksimal setiap dataset skenario 2c.....	48
Gambar 4.19	Grafik evaluasi tertinggi plot <i>silhouette coefficient</i> skema 2	49

DAFTAR LAMPIRAN

Lampiran 1. <i>Centroid</i> akhir dan hasil <i>clustering</i> evaluasi tertinggi dataset Iris ...	53
Lampiran 2. <i>Centroid</i> akhir dan hasil <i>clustering</i> evaluasi tertinggi dataset Wine	58
Lampiran 3. <i>Centroid</i> akhir dan hasil <i>clustering</i> evaluasi tertinggi dataset Glass	60
Lampiran 4. <i>Centroid</i> akhir dan hasil <i>clustering</i> evaluasi tertinggi dataset Ecoli	62
Lampiran 5. <i>Centroid</i> akhir dan hasil <i>clustering</i> evaluasi tertinggi dataset WDBC	65
.....	65
Lampiran 6. Link lengkap data hasil pengujian	69
Lampiran 7. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 1a.	69
Lampiran 8. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 1b	69
Lampiran 9. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 1c.	69
Lampiran 10. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 2a	69
Lampiran 11. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 2b	69
.....	69
Lampiran 12. Link hasil <i>clustering</i> evaluasi maksimal setiap dataset skenario 2c	69