

SKRIPSI



PENERAPAN ALGORITMA GENETIKA DAN *PARTICLE SWARM OPTIMIZATION* (PSO) UNTUK OPTIMASI *K-MEANS* PADA PENGELOMPOKAN PENGGUNA SHOPEE

ANNISYA SAFA KUSYANTI

NIM. 1910511057

**PROGRAM STUDI INFORMATIKA, PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA**

2024

SKRIPSI

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana
Komputer**



**PENERAPAN ALGORITMA GENETIKA DAN *PARTICLE SWARM
OPTIMIZATION* (PSO) UNTUK OPTIMASI *K-MEANS* PADA
PENGELOMPOKAN PENGGUNA SHOPEE**

ANNISYA SAFA KUSYANTI

NIM. 1910511057

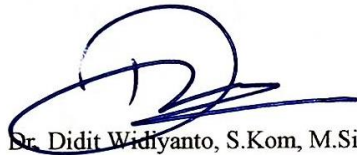
**PROGRAM STUDI INFORMATIKA, PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**

LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Annisya Safa Kusyanti
NIM : 1910511057
Program Studi : S1 Informatika
Judul Tugas Akhir : Penerapan Algoritma Genetika dan *Particle Swarm Optimization* (PSO) untuk Optimasi *K-Means* pada Pengelompokan Pengguna Shopee

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta.



Dr. Didit Widlyanto, S.Kom, M.Si.

Penguji I



Kraugusteeliana, S.Kom., M.Kom., MM.

Penguji II



IIn Ernawati, S.Kom., M.Si.

Pembimbing I



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.

Pembimbing II



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM.

Dekan



Dr. Widya Cholil, M.I.T.

Ketua Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 12 Januari 2024

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Annisya Safa Kusyanti
NIM : 1910511057
Tanggal : 09 Januari 2024
Judul Skripsi : Penerapan Algoritma Genetika dan *Particle Swarm Optimization* (PSO) untuk Optimasi *K-Means* pada Pengelompokan Pengguna Shopee

Bilamana pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 09 Januari 2024

Yang Menyatakan,



Annisya Safa Kusyanti

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Annisya Safa Kusyanti
NIM : 1910511057
Fakultas : Ilmu Komputer
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan karya ilmiah saya kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exchange Royalty Free Right*) untuk dipublikasikan dengan judul:

Penerapan Algoritma Genetika Dan *Particle Swarm Optimization* (PSO) untuk Optimasi *K-Means* pada Pengelompokan Pengguna Shopee

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalis media atau memformatkan, mengelola dalam bentuk pengkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 20 Januari 2024

Yang Menyatakan,



Annisya Safa Kusyanti

KATA PENGANTAR

Puji dan syukur peneliti panjatkan ke hadirat Allah SWT karena atas berkah dan rahmatnya, penulis dapat menyelesaikan penyusunan skripsi dengan judul “Penerapan Algoritma Genetika dan *Particle Swarm Optimization* (PSO) untuk Optimasi *K-Means* pada Pengelompokan Pengguna Shopee”.

Skripsi ini disusun guna memenuhi persyaratan memperoleh gelar sarjana Ilmu Komputer pada Jurusan Informatika di Universitas Pembangunan Nasional “Veteran” Jakarta. Dalam penyusunannya, peneliti mendapat bantuan dari berbagai pihak sehingga skripsi ini dapat diselesaikan dengan baik. Maka pada kesempatan ini, peneliti ingin menyampaikan ucapan terima kasih dan rasa hormat kepada:

1. Bapak, Mama, Kakak, dan Adik-Adik peneliti yang terus memberikan dukungan dan doa kepada peneliti.
2. Ibu Iin Ernawati, S.Kom., M.Si. selaku dosen pembimbing satu skripsi dan dosen pembimbing akademik yang telah membimbing peneliti menyusun skripsi dengan baik.
3. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. selaku dosen pembimbing dua skripsi yang selalu dengan sabar membimbing, memberikan semangat, dan doa kepada peneliti.
4. Ibu Helena Nurramdhani Irmanda, S.Pd, M.Kom. yang telah membantu membimbing peneliti di awal penelitian dan penyusunan skripsi.
5. Ibu Dr. Widya Cholil, M.I.T. sebagai ketua program studi Informatika.
6. Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM. sebagai Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jakarta.
7. Bapak/Ibu Dosen Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jakarta yang telah mengajar dan membagikan ilmunya selama peneliti menjadi mahasiswa di Fakultas Ilmu Komputer.

8. Ajeng, Putri, dan sahabat-sahabat peneliti lain yang telah sangat membantu peneliti selama masa studi dan penelitian di Universitas Pembangunan Nasional “Veteran” Jakarta.

Peneliti menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, peneliti menerima segala bentuk kritik dan saran yang dapat dikirim melalui *e-mail* annisyasafa94@gmail.com agar skripsi ini dapat memberikan sebanyak-banyaknya manfaat dalam pengembangan ilmu pengetahuan.

Jakarta, 08 Januari 2024
Peneliti,

Annisya Safa Kusyanti

**PENERAPAN ALGORITMA GENETIKA DAN *PARTICLE SWARM*
OPTIMIZATION (PSO) UNTUK OPTIMASI *K-MEANS* PADA
PENGELOMPOKAN PENGGUNA SHOPEE**

ANNISYA SAFA KUSYANTI

ABSTRAK

Shopee merupakan salah satu saluran belanja elektronik yang paling banyak digunakan masyarakat Indonesia. Dalam melakukan promosi terhadap barang dagangannya, pelaku usaha di *platform* Shopee dapat menggunakan pengelompokan pada pengguna Shopee untuk memperoleh target pasar. Algoritma pengelompokan yang paling terkenal adalah *K-Means*. Namun *K-Means* memiliki kekurangan yaitu rentan terjebak dalam optima lokal yang disebabkan inisialisasi titik pusat kelompok secara acak. Oleh karena itu, peneliti melakukan proses pengelompokan pada data pengguna Shopee dengan algoritma *K-Means* dan algoritma optimasi Genetika dan PSO untuk menangani kekurangan *K-Means* tersebut. Data yang digunakan peneliti memiliki kategori variabel demografis dan perilaku pengguna. Data didapatkan dari melakukan survei dengan jumlah sampel ditentukan oleh metode *Slovin*. Peneliti membuat 3 (tiga) buah model pengelompokan, yaitu *K-Means* standar, *K-Means* dengan Algoritma Genetika (GA-KMeans), dan *K-Means* dengan PSO (PSO-KMeans) yang kemudian ketiga model tersebut dievaluasi dengan *Silhouette Coefficient* (SC). Hasil penelitian membuktikan bahwa baik GA-KMeans dan PSO-KMeans terbukti dapat mengoptimasi *K-Means* dengan dibuktikan dari meningkatnya nilai SC. Dengan jumlah kelompok $k = 3$, GA-KMeans memperoleh nilai SC = 0,4281 dan PSO-KMeans sebesar SC = 0,4075. Sedangkan nilai SC *K-Means* standar = 0,2725.

Kata kunci: pengelompokan, *K-Means*, Algoritma Genetika, PSO, pengguna Shopee

**IMPLEMENTATION OF GENETIC ALGORITHM AND PARTICLE
SWARM OPTIMIZATION (PSO) FOR K-MEANS OPTIMIZATION ON
SHOPEE USERS CLUSTERING**

ANNISYA SAFA KUSYANTI

ABSTRACT

Shopee is one of the most widely used electronic shopping channels in Indonesia. In promoting their products, businesses on the Shopee platform can use clustering on Shopee users to obtain their target market. The most well-known clustering algorithm is K-Means. However, K-Means has a disadvantage in that it is susceptible to being trapped in local optima caused by the random initialization of cluster center points. Therefore, researchers have conducted clustering on Shopee user data using the K-Means algorithm and Genetic and PSO optimization algorithms to address this K-Means limitation. The data used by the researchers had demographic and user behavior variable categories, obtained from a survey with a sample size determined by the Slovin method. The researchers created three clustering models, namely standard K-Means, K-Means with Genetic Algorithm (GA-KMeans), and K-Means with PSO (PSO-KMeans), which were then evaluated using the Silhouette Coefficient (SC). The research results prove that both GA-KMeans and PSO-KMeans can optimize K-Means. With several clusters $k = 3$, GA-KMeans obtained an SC value of 0.4281 and PSO-KMeans obtained an SC value of 0.4075. Meanwhile, the SC value of standard K-Means was 0.2725.

Keywords: *clustering, K-Means, Genetic Algorithm, PSO, Shopee users*

DAFTAR ISI

HALAMAN AWAL	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	v
KATA PENGANTAR	vi
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	4
1.5 Ruang Lingkup	4
1.6 Luaran yang Diharapkan	5
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 <i>Customer Segmentation</i>	7
2.2 Shopee	7
2.3 Machine Learning	8
2.4 <i>Clustering</i>	9
2.5 <i>K-Means Clustering</i>	9
2.6 Algoritma Genetika	11
2.6.1 Seleksi Individu	13
2.6.2 Proses Persilangan (<i>Crossover</i>)	14

2.6.3	Proses Mutasi	15
2.6.4	<i>Elitism</i>	16
2.7	<i>Particle Swarm Optimization</i> (PSO).....	16
2.8	Praproses Data	18
2.8.1	<i>Data Cleaning</i>	19
2.8.2	<i>Label Encoding</i>	19
2.9	Silhouette Coefficient	20
2.10	Metode <i>Slovin</i>	20
2.11	Penelitian Terdahulu	21
BAB III METODE PENELITIAN.....		27
3.1	Kerangka Berpikir.....	27
3.1.1	Identifikasi Masalah	27
3.1.2	Studi Literatur	28
3.1.3	Penyusunan Kuesioner	28
3.1.4	Pengumpulan Data	30
3.1.5	Praproses Data.....	31
3.1.6	Pemodelan Data	32
3.1.7	Evaluasi Model.....	33
3.1.8	Uji Coba dan Perbandingan Model Pengelompokan	34
3.1.9	Analisis Hasil Pengelompokan	34
3.2	Sarana Pendukung.....	34
3.3	Waktu Kegiatan	35
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Akuisisi Data.....	37
4.2	Praproses Data	39
4.2.1	<i>Data Cleaning</i>	39
4.2.2	<i>Label Encoding</i>	39
4.3	Pembuatan Model Pengelompokan dengan <i>K-Means</i> Standar	43
4.4	Pembuatan Model Pengelompokan dengan <i>K-Means</i> dan Optimasi Algoritma Genetik (GA-KMeans)	49

4.5	Pembuatan Model Pengelompokan dengan <i>K-Means</i> dan Optimasi <i>Particle Swarm Optimization</i> (PSO).....	67
4.6	Evaluasi Hasil Pengelompokan.....	86
4.7	Uji Coba dan Perbandingan Model Pengelompokan.....	93
4.7.1	Uji Coba <i>K-Means</i> Standar	93
4.7.2	Uji Coba GA-KMeans.....	94
4.7.3	Uji Coba PSO-KMeans	96
4.7.4	Perbandingan Algoritma	98
4.8	Analisis Hasil Pengelompokan	101
BAB V KESIMPULAN DAN SARAN.....		111
5.1	Kesimpulan	111
5.2	Saran	112
DAFTAR PUSTAKA		113
RIWAYAT HIDUP.....		119
LAMPIRAN.....		120
	Lampiran 1. Hasil Turnitin.....	120

DAFTAR GAMBAR

Gambar 2.1. Jenis machine learning menurut Bell (2022).....	8
Gambar 2.2. Flowchart Algoritma K-Means Clustering (Wu, 2012).	10
Gambar 2.3. Visualisasi Istilah-Istilah GA (Ahmed et al., 2020).	11
Gambar 2.4. Flowchart Proses Algoritma Genetika (Kramer, 2017).	12
Gambar 2.5. Proses Seleksi pada GA 5 (Gurpinar et al., 2021).	13
Gambar 2.6. Proses Crossover pada GA (Mirjalili, 2019).	14
Gambar 2.7. Proses Mutasi pada GA (Mirjalili, 2019).	15
Gambar 2.8. Flowchart proses Algoritma PSO (Yang, 2014).	17
Gambar 3.1. Flowchart Tahapan Penelitian.	27
Gambar 4.1. Kode Program Algoritma K-Means.	49
Gambar 4.2. Proses Seleksi Iterasi 1.	56
Gambar 4.3. Proses Crossover Iterasi 1.	58
Gambar 4.4. Proses Mutasi Iterasi 1.	59
Gambar 4.5. Kode Program Algoritma Genetika.....	67
Gambar 4.6. Kode Program Algoritma PSO.....	85
Gambar 4.7. Visualisasi Perbandingan Evaluasi Model Pengelompokan.	99
Gambar 4.8. Visualisasi Perbandingan Waktu Komputasi Tiap Model Pengelompokan.....	100
Gambar 4.9. Distribusi Anggota Cluster GA-KMeans.	103
Gambar 4.10. Distribusi Variabel <i>Cluster</i> 0 untuk model GA-Kmeans.	104
Gambar 4.11. Distribusi Variabel Cluster 1 untuk model GA-Kmeans.	106
Gambar 4.12. Distribusi Variabel Cluster 2 untuk model GA-Kmeans.	108

DAFTAR TABEL

Tabel 2.1. Studi Literatur.	22
Tabel 3.1. Rincian Variabel Kategori Demografis.....	29
Tabel 3.2. Rincian Variabel Kategori Perilaku.	29
Tabel 3.3. Jadwal Penelitian.....	35
Tabel 4.1. Data Pengguna Shopee.....	38
Tabel 4.2. Hasil Label Encoding.....	39
Tabel 4.3. Data Bersih Pengguna Shopee.	41
Tabel 4.4. Data Sampel Pengguna Shopee.....	42
Tabel 4.5. Inisialisasi Centroid K-Means Standar.	43
Tabel 4.6. Euclidean Distance Iterasi 1 pada K-Means Standar.	44
Tabel 4.7. Centroid Iterasi 1 pada K-Means Standar.	47
Tabel 4.8. Proses Perubahan Nilai Centroid pada K-Means Standar.....	48
Tabel 4.9. Parameter Algoritma Genetika.....	50
Tabel 4.10. Inisialisasi Populasi Algoritma Genetika.	51
Tabel 4.11. Euclidean Distance Individu Pertama Iterasi 1.	52
Tabel 4.12. Nilai Fitness Tiap Individu pada Iterasi 1.	53
Tabel 4.13. Kumulatif dan Rentang untuk Proses Seleksi Iterasi 1.	55
Tabel 4.14. Bilangan Acak dan Individu Terpilih Iterasi 1.....	55
Tabel 4.15. Populasi Setelah Proses Evolusi Iterasi 1.	60
Tabel 4.16. Populasi Setelah Elitism.....	61
Tabel 4.17. Centroid Hasil Proses Algoritma Genetika.	61
Tabel 4.18. Euclidean distance Iterasi 1 pada GA-KMeans.	62
Tabel 4.19. Perubahan Nilai Centroid pada GA-KMeans.....	65
Tabel 4.20. Parameter Algoritma PSO.....	67
Tabel 4.21. Titik Pusat Cluster untuk Algoritma PSO.....	69
Tabel 4.22. Euclidean Distance Partikel Pertama Iterasi t.	70
Tabel 4.23. Nilai Fitness Tiap Partikel pada Iterasi t.	71
Tabel 4.24. Inisialisasi Partikel Iterasi t.	72
Tabel 4.25. Inisialisasi Kecepatan Iterasi t.....	73

Tabel 4.26. Inisialisasi Pbest Iterasi t.....	73
Tabel 4.27. Inisialisasi Gbest Iterasi t.....	74
Tabel 4.28. Hasil Perhitungan Perbarui Kecepatan Iterasi 1.....	75
Tabel 4.29. Hasil Konversi Perbarui Kecepatan Iterasi 1.....	76
Tabel 4.30. Hasil Perhitungan Perbarui Posisi Partikel Iterasi 1.....	77
Tabel 4.31. Nilai Fitness Tiap Partikel pada Iterasi t+1.....	77
Tabel 4.32. Hasil Perbarui Pbest Iterasi 1.....	78
Tabel 4.33. Hasil Perbarui Gbest Iterasi 1.....	79
Tabel 4.34. Proses Perubahan Nilai Gbest dari Algoritma PSO.....	79
Tabel 4.35. Centroid Hasil Proses Algoritma PSO.....	80
Tabel 4.36. Euclidean Distance Iterasi 1 pada PSO-KMeans.....	81
Tabel 4.37. Perubahan Nilai Centroid pada PSO-KMeans.....	84
Tabel 4.38. Euclidean distance Antar Variabel.....	87
Tabel 4.39. Perhitungan Silhouette Coefficient untuk K-Means Standar Data Sampel.....	89
Tabel 4.40. Perhitungan Silhouette Coefficient untuk GA-KMeans Data Sampel.....	91
Tabel 4.41. Perhitungan Silhouette Coefficient untuk PSO-KMeans Data Sampel.....	92
Tabel 4.42. Uji Coba Jumlah Cluster K-Means Standar.....	93
Tabel 4.43. Uji Coba Jumlah Individu dalam Populasi.....	94
Tabel 4.44. Uji Coba Jumlah Generasi.....	95
Tabel 4.45. Uji Coba Jumlah Cluster GA-KMeans.....	95
Tabel 4.46. Uji Coba Jumlah Partikel.....	96
Tabel 4.47. Uji Coba Jumlah Iterasi.....	97
Tabel 4.48. Uji Coba Jumlah Cluster PSO-KMeans.....	97
Tabel 4.49. Perbandingan Nilai Silhouette Tiap Model Pengelompokan.....	98
Tabel 4.50. Hasil Pengelompokan Data Utuh GA-KMeans dengan Jumlah Cluster 3.....	102