

OPTIMASI ALGORITMA GENETIKA DALAM INISIALISASI PUSAT GEROMBOL *K-PROTOTYPES* (STUDI KASUS: MATA KULIAH S1 IPB UNIVERSITY)

AVREL CHESIA BERBINA



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Optimasi Algoritma Genetika dalam Inisialisasi Pusat Gerombol *K-Prototypes* (Studi Kasus: Mata Kuliah S1 IPB University)” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Penulis menggunakan ChatGPT dan Claude untuk membantu perbaikan tata bahasa, penyusunan kalimat, serta pencarian dan pengembangan *syntax* pemrograman guna mendukung analisis data. Penulis meninjau dan menyunting seluruh konten serta kode yang dihasilkan, dan bertanggung jawab penuh atas isi karya tugas akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Avrel Chesia Berbina
G1401221050



ABSTRAK

AVREL CHESIA BERBINA. Optimasi Algoritma Genetika dalam Inisialisasi Pusat Gerombol K-Prototypes (Studi Kasus: Mata Kuliah S1 IPB University). Dibimbing oleh ERFIANI dan UTAMI DYAH SYAFITRI.

Sensitivitas algoritme *K-Prototypes* terhadap pemilihan *centroid* awal sering menghambat penggerombolan data bertipe campuran. Inisialisasi secara acak cenderung menghasilkan solusi optimum lokal sehingga hasil penggerombolan menjadi kurang stabil pada setiap pengulangan. Penelitian ini bertujuan mengoptimalkan pemilihan *centroid* awal pada algoritme *K-Prototypes* menggunakan algoritme genetika. Metode tersebut diterapkan pada data mata kuliah program S1 IPB University sebanyak 1.364 mata kuliah yang terdiri atas tujuh peubah kategorik dan sepuluh peubah numerik. Keباikan gerombol yang dihasilkan dinilai berdasarkan nilai *within-cluster sum of squares* (WCSS), dengan nilai yang semakin rendah menunjukkan gerombol yang semakin kompak. Dua belas skenario algoritma genetika dievaluasi berdasarkan variasi ukuran populasi, jenis *crossover*, dan jumlah generasi maksimum, dengan kinerja diukur menggunakan WCSS dan waktu komputasi. Skenario dengan ukuran populasi 100, one-point crossover, dan 100 generasi dipilih sebagai konfigurasi terbaik karena secara konsisten menghasilkan nilai WCSS terendah dengan waktu komputasi paling efisien. Kombinasi ini menghasilkan WCSS sebesar 24.021,26 pada lima gerombol. Nilai WCSS ini lebih rendah 5,18% dibandingkan inisialisasi acak. Kelima gerombol tersebut terpisah berdasarkan tipe mata kuliah, jumlah kelas paralel, keberadaan praktikum, keberadaan responsi, kapasitas peserta, dan durasi pertemuan. Algoritme genetika terbukti memperbaiki stabilitas proses penggerombolan *K-Prototypes*. Kualitas hasil gerombol turut meningkat dibandingkan dengan pendekatan inisialisasi acak.

Kata kunci: algoritme genetika, inisialisasi centroid, K-Prototypes, penggerombolan.



ABSTRACT

AVREL CHESIA BERBINA. Optimization of K-Prototypes Cluster Centroid Initialization Using a Genetic Algorithm (Case Study: Undergraduate Course Data at IPB University. Supervised by ERFIANI and UTAMI DYAH SYAFITRI.

The sensitivity of the K-Prototypes algorithm to initial centroid selection often hinders clustering of mixed-type data. Random initialization tends to produce local optima, resulting in less stable clustering outcomes across repetitions. This study aims to optimize the selection of initial centroids in the K-Prototypes algorithm using a genetic algorithm. The method was applied to course data from undergraduate programs at IPB University, comprising 1,364 courses with seven categorical and ten numerical variables. The quality of the resulting clusters was evaluated based on the within-cluster sum of squares (WCSS), where lower values indicate more compact clusters. Twelve genetic algorithm scenarios were evaluated based on population size, crossover type, and maximum number of generations, with performance measured using WCSS and computational time. The scenario with a population size of 100, one-point crossover, and 100 generations was selected as the best configuration because it consistently produced the lowest WCSS with the most efficient computational time. This combination yielded a WCSS of 24,021.26 for five clusters, which was 5.18% lower than that obtained using random initialization. The five clusters were distinguished by course type, number of parallel classes, availability of practical sessions, availability of tutorials, participant capacity, and meeting duration. The genetic algorithm improved the stability and cluster quality of K-Prototypes compared with random initialization.

Keywords: centroid initialization, clustering, genetic algorithm, K-prototypes.



@Hak cipta milik IPB University

IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

OPTIMASI ALGORITMA GENETIKA DALAM INISIALISASI PUSAT GEROMBOL *K-PROTOTYPES* (STUDI KASUS: MATA KULIAH S1 IPB UNIVERSITY)

AVREL CHESIA BERBINA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Skripsi:
Akbar Rizki, S.Stat., M.Si



IPB University
— Bogor Indonesia —

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



Judul Skripsi : Optimasi Algoritma Genetika dalam Inisialisasi Pusat Gerombol
K-Prototypes (Studi Kasus: Mata Kuliah S1 IPB University)

Nama : Avrel Chesia Berbina
NIM : G1401221050

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Ir. Erfiani, M.Si.

Pembimbing 2:

Dr. Utami Dyah Syafitri, S.Si., M.Si.

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si, M.Si.
NIP 197804112005011002

Tanggal Ujian:
23 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia-Nya sehingga karya ilmiah yang dilaksanakan sejak bulan Januari 2026 sampai bulan Juli 2026 ini dengan judul “Optimasi Algoritma Genetika dalam Inisialisasi Pusat Gerombol *K-Prototypes* (Studi Kasus: Mata Kuliah S1 IPB University)” bisa diselesaikan. Pada proses pembuatan karya ilmiah ini, penulis banyak sekali mendapatkan doa, bimbingan, serta dukungan dari berbagai pihak.

Penulis mengucapkan terima kasih kepada Ibu Dr. Ir. Erfiani, M.Si. dan Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. selaku dosen pembimbing atas arahan dan masukan selama penyusunan karya ilmiah ini. Terima kasih juga penulis sampaikan kepada Ibu Akbar Rizki, S.Stat., M.Si. selaku dosen penguji sekaligus moderator seminar hasil, dan Ibu Dr. Yenni Angraini, S.Si., M.Si. selaku moderator kolokium atas saran perbaikan karya ilmiah ini. Ucapan terima kasih disampaikan kepada Bapak Pharmasi Leo, Ibu Aria Christy, Aurey Fabio Seymsalio, Iting, serta seluruh keluarga besar atas doa, dukungan, dan kasih sayang dalam setiap perjalanan penulis. Terima kasih juga penulis sampaikan kepada Jasmita, Yumna, Naiya, Keyzha, Aura, Farid, dan Cecil selaku sahabat yang memberi warna selama empat tahun perkuliahan, serta kepada Sandra, Oci, dan Kak Fikri atas bantuannya dalam penyusunan karya ilmiah ini. Terima kasih juga disampaikan kepada teman-teman Statistika 59 Marinestic atas kebersamaan dan dukungannya selama masa perkuliahan.

Bogor, Juli 2026

Avrel Chesia Berbina



DAFTAR ISI

DAFTAR TABEL	vii
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	vii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Analisis Gerombol	3
2.2 Algoritma <i>K-Means</i>	3
2.3 Algoritma <i>K-Prototypes</i>	3
2.4 Algoritma Genetika	5
2.5 Jumlah Gerombol Optimal	8
2.6 ANOVA	8
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis Data	9
IV HASIL DAN PEMBAHASAN	13
4.1 Analisis Eksplorasi Data	13
4.2 Evaluasi Algoritma Genetika dengan 12 Skenario	14
4.3 Pemilihan Kombinasi Terbaik	17
4.4 Pusat Awal Gerombol Acak dan Optimal	17
4.5 Perbandingan Metode Inisiasi Pusat Awal Gerombol	18
4.6 Perbandingan Hasil Penggerombolan	19
V SIMPULAN DAN SARAN	28
5.1 Simpulan	28
5.2 Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN	30
RIWAYAT HIDUP	42

DAFTAR TABEL

1	Peubah yang digunakan	9
2	Skenario algoritma genetika	11
3	P-Value ANOVA	16
4	Pusat awal gerombol acak	17
5	Pusat awal gerombol optimal	17
6	Perbandingan total iterasi dan waktu komputasi	18
7	Perbandingan rata-rata total jarak	18
8	Karakteristik peubah kategorik pada hasil penggerombolan dengan pusat awal acak	20
9	Karakteristik peubah numerik hasil penggerombolan dengan pusat awal acak	21
10	Sebaran unit pengampu pada hasil penggerombolan dengan pusat awal acak	22
11	Karakteristik peubah kategorik pada hasil penggerombolan dengan pusat awal optimal	24
12	Karakteristik peubah numerik pada hasil penggerombolan dengan pusat awal optimal	25
13	Sebaran unit pengampu pada hasil penggerombolan dengan pusat awal acak	26

DAFTAR GAMBAR

1	<i>One-point crossover</i>	6
2	<i>Two-points crossover</i>	7
3	Sebaran kategori mata kuliah	13
4	Distribusi mata kuliah berdasarkan jenis perkuliahan	13
5	Sebaran mata kuliah berdasarkan peubah numerik	14
6	Perubahan total jarak (a) pada pusat gerombol acak dan (b) pada pusat awal gerombol optimal	19

DAFTAR LAMPIRAN

1	Jenis kategori peubah kategorik	31
2	Fungsi <i>two-points crossover</i> pada software R	32
3	Rataan total jarak dan simpangan baku	33
4	Hasil uji ANOVA untuk $k = 3$ sampai $k=6$	34
5	Uji Asumsi	34
6	Rataan waktu komputasi dan simpangan baku	35
7	Sebaran (X1) pada setiap gerombol dengan <i>centroid</i> acak	36
8	Sebaran (X1) pada setiap gerombol dengan <i>centroid</i> optimal	39