

- 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.

PENERAPAN ALGORITMA GENETIKA DALAM PENENTUAN PUSAT GEROMBOL K-PROTOTYPES (STUDI KASUS: DATA MATA KULIAH JENJANG S1 IPB UNIVERSITY)

ALFIKRI IHSAN



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



IPB University

IPB University
— Bogor Indonesia —



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penerapan Algoritma Genetika dalam Penentuan Pusat Awal Gerombol *K-Prototypes* (Studi Kasus: Data Mata Kuliah Jenjang 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.

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

Bogor, Juni 2025

Alfikri Ihsan
G1401211058



ABSTRAK

ALFIKRI IHSAN. Penerapan Algoritma Genetika dalam Penentuan Pusat Awal Gerombol *K-Prototypes* (Studi Kasus: Data Mata Kuliah Jenjang S1 IPB University). Dibimbing oleh UTAMI DYAH SYAFITRI dan MUHAMMAD NUR AIDI.

Analisis gerombol merupakan salah satu metode dalam analisis data yang digunakan untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristiknya. Pada umumnya, analisis gerombol banyak digunakan dalam pengelompokan data dengan tipe numerik. Pada kenyataannya, banyak data merupakan gabungan dari data bertipe numerik dan juga kategorik. Algoritma *K-Prototypes* adalah salah satu metode dalam analisis gerombol yang dapat menangani permasalahan data campuran. Algoritma ini bekerja dengan cara mengintegrasikan perhitungan jarak pada *K-Means* dan *K-Modes* secara bersamaan. Hasil penggerombolan dengan algoritma *K-Prototypes* cenderung memperoleh solusi yang optimum lokal. Oleh karena itu, pada penelitian dilakukan penggerombolan dengan algoritma *K-Prototypes* dan mengombinasikannya dengan algoritma genetika untuk menentukan pusat awal gerombol yang optimal. Pada penelitian ini penggerombolan akan dilakukan untuk mengelompokkan data mata kuliah program S1 IPB University, yang terdiri dari peubah numerik dan peubah kategorik. Hasil dari penelitian ini diperoleh bahwa penerapan algoritma genetika dalam penentuan pusat awal gerombol mampu menghasilkan solusi yang lebih optimal. Hasil dari penggerombolan menunjukkan bahwa terdapat perbedaan dari karakteristik masing-masing mata kuliah dari jenis pembelajaran, durasi pelaksanaan dan juga kapasitas masing-masing mata kuliah.

Kata kunci: algoritma genetika, analisis gerombol, data campuran, *k-prototypes*

ABSTRACT

ALFIKRI IHSAN. Application of Genetic Algorithm in Determining Cluster Centers of K-Prototypes (Case Study: Undergraduate Course Data at IPB University). Supervised by UTAMI DYAH SYAFITRI and MUHAMMAD NUR AIDI.

Cluster analysis is data analysis methods used to group data based on similarity in characteristics. Typically, cluster analysis is widely used for grouping numerical data. However, in reality, many datasets consist of a combination of both numerical and categorical data. The K-Prototypes algorithm is a clustering technique that can handle mixed-type data, combining both numerical and categorical variables. It works by integrating the distance measures from K-Means and K-Modes simultaneously. However, the solutions generated by the K-Prototypes algorithm tend to result in local optimum solutions. Therefore, this study aims to use the K-Prototypes algorithm optimized with a Genetic Algorithm (GA) to determine the cluster center points, ensuring a global optimum solution. This approach will be applied to cluster the undergraduate course data at IPB University, which includes both numerical and categorical variables. The results of this study indicate that the genetic algorithm can optimize the performance of K-Prototypes by determining optimal initial cluster centers. The clustering results indicate clear variation in the course structure at the university, related to the specific needs of each faculty, the type of learning (practical vs. theoretical), as well as the intensity of the courses (long vs. short duration).

Keywords: cluster analysis, genetic algorithm, k-prototypes, mixed data



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 2025
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.



PENERAPAN ALGORITMA GENETIKA DALAM PENENTUAN PUSAT GEROMBOL *K-PROTOTYPES* (STUDI KASUS: DATA MATA KULIAH JENJANG S1 IPB UNIVERSITY)

ALFIKRI IHSAN

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

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik 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.



Judul Skripsi : Penerapan Algoritma Genetika dalam Penentuan Pusat Awal
Gerombol *K-Prototypes* (Studi Kasus: Data Mata Kuliah Jenjang
S1 IPB University)

Nama : Alfikri Ihsan
NIM : G1401211058

@Hak cipta milik IPB University

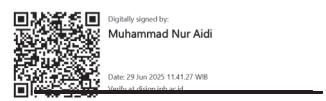
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan 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.

Disetujui oleh

Pembimbing 1:
Dr. Utami Dyah Syafitri, S.Si., M.Si.



Pembimbing 2:
Prof. Dr. Ir. Muhammad Nur Aidi, M.S.



Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002



Tanggal Ujian:
10 Juni 2025

Tanggal Lulus:



@Hak cipta milik 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.

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah dengan judul “Penerapan Algoritma Genetika dalam Penentuan Pusat Awal Gerombol *K-Prototypes* (Studi Kasus: Data Mata Kuliah Jenjang S1 IPB University)” berhasil diselesaikan.

Terima kasih penulis ucapkan kepada semua pihak yang telah berperan dalam membantu penulis untuk menyusun dan menyelesaikan karya ilmiah ini, diantaranya:

1. Ibu Desdefira selaku ibu penulis, serta keluarga besar penulis yang tidak pernah henti mendoakan dan memberikan dukungan kepada penulis.
2. Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. dan Bapak Prof. Dr. Ir. Muhammad Nur Aidi, M.S. selaku dosen pembimbing yang banyak memberikan arahan dan masukan dalam penyelesaian karya ilmiah ini.
3. Bapak Ir. Mohammad Masjkur, M.S. selaku dosen penguji sidang skripsi yang juga ikut memberikan masukan dan saran untuk penelitian ini.
4. Bapak Dr. Mochamad Farit Afendi, S.Si., M.Si. dan ibu Dr. Yenni Angraini, S.Si., M.Si. selaku moderator kolokium dan seminar hasil yang juga telah memberikan saran dan masukan dalam perbaikan karya ilmiah ini.
5. Keluarga besar Program Studi Statistika dan Sains Data khususnya teman-teman seperjuangan Statistika 58 yang telah mengisi dan mewarnai hari-hari penulis selama penyusunan karya ilmiah ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juni 2025

Alfikri Ihsan



DAFTAR ISI

DAFTAR TABEL	iii
DAFTAR GAMBAR	iii
DAFTAR LAMPIRAN	iii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Analisis Gerombol	3
2.2 <i>K-Prototypes</i>	3
2.3 Algoritma Genetika	4
2.4 Penentuan Jumlah Gerombol Optimal	5
III METODE	7
3.1 Data	7
3.2 Prosedur Analisis Data	7
IV HASIL DAN PEMBAHASAN	9
4.1 Analisis Eksplorasi Data	9
4.2 Penerapan Algoritma Genetika pada <i>K-Prototypes</i>	10
4.3 Perbandingan Metode Inisialisasi Pusat Awal Gerombol	11
4.4 Jumlah Gerombol Optimal	12
4.5 Hasil Penggerombolan dengan Pusat Awal Acak	13
4.6 Hasil Penggerombolan dengan Pusat Awal Optimal	16
V SIMPULAN DAN SARAN	20
5.1 Simpulan	20
5.2 Saran	20
DAFTAR PUSTAKA	21
LAMPIRAN	23
RIWAYAT HIDUP	29

DAFTAR TABEL

1	Peubah yang digunakan	7
2	Distribusi Mata Kuliah Berdasarkan Jenis Pelaksanaan	9
3	Pusat Awal Gerombol Optimal	11
4	Perbandingan Total Jarak	11
5	Sebaran Program Studi pada Hasil Penggerombolan dengan Pusat Awal Acak	13
6	Karakteristik Hasil Penggerombolan dengan Pusat Awal Acak	14
7	Sebaran Program Studi pada Hasil Penggerombolan dengan Pusat Awal Optimal	16
8	Karakteristik Hasil Penggerombolan dengan Pusat Awal Optimal	16

DAFTAR GAMBAR

1	Sebaran kategori mata kuliah	9
2	Sebaran mata kuliah (a) <i>enrichment course</i> dan (b) berdasarkan semester	9
3	Sebaran mata kuliah berdasarkan peubah numerik	10
4	Perubahan nilai <i>fitness</i> untuk $k = 2$	11
5	Perubahan nilai <i>fitness</i> untuk $k = 5$	11
6	Sebaran total jarak untuk setiap nilai k dengan pusat awal acak	13
7	Perubahan total jarak (a) pada pusat gerombol acak dan (b) pusat gerombol optimal	14

DAFTAR LAMPIRAN

1	Lampiran 1 Jenis kategori untuk setiap peubah kategorik	24
2	Lampiran 2 Modifikasi fungsi <i>K-Prototypes</i> pada <i>software R</i>	25
3	Lampiran 3 Fungsi Algoritma genetika dalam penentuan pusat awal gerombol optimal	28