

PENERAPAN ALGORITME *K-PROTOTYPES* UNTUK SEGMENTASI MAHASISWA BARU PROGRAM SARJANA IPB UNIVERSITY TAHUN 2024 BERDASARKAN KARAKTERISTIKNYA

OKTAVIA GALIH PRATIWI



**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.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penerapan Algoritme *K-Prototypes* untuk Segmentasi Mahasiswa Baru Program Sarjana IPB University Tahun 2024 Berdasarkan Karakteristiknya” 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, Juli 2025

Oktavia Galih Pratiwi
G1401211066

- 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.



ABSTRAK

OKTAVIA GALIH PRATIWI. Penerapan Algoritme *K-Prototypes* untuk Segmentasi Mahasiswa Baru Program Sarjana IPB University Tahun 2024 Berdasarkan Karakteristiknya. Dibimbing oleh AGUS MOHAMAD SOLEH dan GERRY ALFA DITO.

IPB University setiap tahun menerima mahasiswa baru dari latar belakang beragam sehingga muncul tantangan dalam penyediaan layanan yang adil dan tepat sasaran. Salah satu pendekatan untuk mengatasi tantangan tersebut adalah segmentasi mahasiswa menggunakan algoritme *K-Prototypes*, yang dirancang untuk menangani data campuran dengan menggabungkan prinsip *K-Means* (numerik) dan *K-Modes* (kategorik). Penelitian ini menerapkan algoritme *K-Prototypes* untuk segmentasi mahasiswa baru program sarjana IPB University tahun 2024 dan mengidentifikasi karakteristik mahasiswa dari hasil penggerombolan yang optimal. Data mencakup 4723 mahasiswa dengan empat peubah numerik dan sebelas peubah kategorik. Proses penggerombolan dilakukan setelah praproses data, standarisasi, penyederhanaan kategori peubah, dan dievaluasi menggunakan rasio S_W/S_B minimum. Hasil penelitian menunjukkan jumlah gerombol optimal adalah lima dengan rasio S_W/S_B sebesar 0,121. Profil gerombol mencerminkan keragaman kondisi sosial ekonomi mahasiswa mulai dari kelompok rentan hingga kelompok paling mapan. Segmentasi ini diharapkan menjadi masukan strategis dalam penetapan kebijakan UKT, bantuan keuangan, serta strategi penerimaan mahasiswa baru di IPB University.

Kata kunci: data campuran, IPB University, *K-Prototypes*, mahasiswa baru, segmentasi.



ABSTRACT

OKTAVIA GALIH PRATIWI. Implementation of the K-Prototypes Algorithm for Segmenting New Undergraduate Students at IPB University in 2024 Based on Characteristics. Supervised by AGUS MOHAMAD SOLEH and GERRY ALFA DITO.

IPB University annually admits new undergraduate students from diverse backgrounds, posing challenges in delivering fair and targeted services. One approach to overcoming this issue is student clustering using the K-Prototypes algorithm, which is designed to handle mixed-type data by combining the principles of K-Means (for numerical data) and K-Modes (for categorical data). This study applies the K-Prototypes algorithm to segment new undergraduate students at IPB University in 2024 and identifies the characteristics of the optimal clustering results. The dataset comprises 4,723 students with four numerical variables and eleven categorical variables. Clustering was performed following data preprocessing, variable recoding, and standardization, with evaluation using the minimum S_W/S_B ratio. The study results indicate that the optimal number of clusters is five, indicated by the minimum S_W/S_B ratio of 0,121. The cluster profiles reflect the diversity of socio-economic conditions, ranging from vulnerable groups to the most affluent. This segmentation is expected to serve as strategic input for setting tuition fee (UKT) policies, financial aid programs, and new student admission strategies at IPB University.

Keywords: IPB University, K-Prototypes, mixed type data, new students, segmentation.



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 ALGORITME *K-PROTOTYPES* UNTUK SEGMENTASI
MAHASISWA BARU PROGRAM SARJANA IPB UNIVERSITY
TAHUN 2024 BERDASARKAN KARAKTERISTIKNYA**

@Hak cipta milik IPB University

OKTAVIA GALIH PRATIWI

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

IPB University

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

- 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 Algoritme *K-Prototypes* untuk Segmentasi Mahasiswa Baru Program Sarjana IPB University Tahun 2024 Berdasarkan Karakteristiknya

Nama : Oktavia Galih Pratiwi
NIM : G1401211066

Disetujui oleh

Pembimbing 1:
Dr. Agus Mohamad Soleh S.Si., M.T.



Pembimbing 2:
Gerry Alfa Dito, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002



Tanggal Ujian:
1 Juli 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 limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan karya ilmiah yang berjudul “Penerapan Algoritme *K-Prototypes* untuk Segmentasi Mahasiswa Baru Program Sarjana IPB University Tahun 2024 Berdasarkan Karakteristiknya”. Penyusunan karya ilmiah ini merupakan proses yang tidak terlepas dari dukungan, bimbingan, arahan, serta motivasi dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis menyampaikan ucapan terima kasih secara khusus kepada:

1. Bapak Dr. Agus Mohamad Soleh S.Si., M.T. dan Bapak Gerry Alfa Dito, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan arahan dan dukungan selama proses penyusunan karya ilmiah ini.
2. Ibu Akbar Rizki, S.Stat., M.Si. selaku dosen penguji yang telah memberikan saran dan masukan dalam penyelesaian karya ilmiah ini.
3. Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si., Bapak Windra Priawandiputra, S.Si., M.Si., Ph.D., Ibu Ribka Puji Raspati, S.Pi, dan Mba Nabila selaku tim Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPPMB) IPB University atas dukungan dan bantuan yang telah diberikan dalam perizinan dan penyediaan data penelitian.
4. Seluruh dosen dan staf Program Studi Statistika dan Sains Data IPB University yang telah memberikan ilmu dan dukungan administratif selama masa perkuliahan.
5. Bapak Kamidi dan Ibu Mujiati, serta kakak dan adik penulis, yaitu Mayang Yulisyaroh dan Dewi Anjani, beserta seluruh keluarga besar yang selalu memberikan doa, dukungan, dan nasihat dalam setiap langkah perjalanan penulis.
6. Kak Nabila dan Kak Ervina selaku kakak tingkat yang telah banyak membantu penulis melalui diskusi, berbagi pengalaman, dan memberikan masukan selama proses penyusunan karya ilmiah.
7. Ina Marina, teman terdekat penulis sejak masa sekolah dasar yang selalu percaya tanpa ragu dan hadir dalam setiap tahap kehidupan penulis.
8. Farhan, Vita, Farrel, Farras, Fajar, Much, dan Adisti, rekan seperjuangan di bangku perkuliahan sekaligus teman satu kos yang telah banyak membantu dan menjadi bagian penting dalam perjalanan studi penulis.
9. Hanafi, Bayu, dan Manda selaku rekan-rekan penulis dari berbagai jenjang pendidikan yang senantiasa kebersamaian penulis.
10. Mutiara, Zulfa, Dhiya, Syifa, dan Diva, serta teman-teman Gamavectra Statistika 58 yang tidak dapat disebutkan satu per satu atas kebersamaan, semangat, dan dukungannya selama masa perkuliahan.

Semoga karya ilmiah ini dapat memberikan kontribusi positif bagi pihak yang membutuhkan dan membawa nilai tambah bagi perkembangan ilmu pengetahuan.

Bogor, Juli 2025

Oktavia Galih Pratiwi



@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.

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Analisis Gerombol	3
2.2 Ukuran Kemiripan	3
2.3 Algoritme <i>K-Prototypes</i>	4
2.4 Evaluasi Hasil Gerombol	6
2.5 <i>Chernoff Faces</i>	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis	9
IV HASIL DAN PEMBAHASAN	11
4.1 Praproses Data	11
4.2 Eksplorasi Data	13
4.3 Analisis Gerombol dengan Algoritme <i>K-Prototypes</i>	20
4.4 Karakteristik Gerombol Berdasarkan Peubah Numerik	22
4.5 Karakteristik Gerombol Berdasarkan Peubah Kategorik	27
4.6 Profil Segmentasi Mahasiswa Baru Berdasarkan Karakteristik	34
V SIMPULAN	38
DAFTAR PUSTAKA	39
LAMPIRAN	42
RIWAYAT HIDUP	63



DAFTAR TABEL

1	Deskripsi parameter <i>chernoff faces</i> pada paket R <i>aplpack</i>	8
2	Peubah-peubah yang digunakan dalam penelitian	9
3	Perbandingan jumlah kategori awal dan akhir pada peubah kategorik	13
4	Nilai rasio S_W dan S_B	22
5	Banyaknya anggota gerombol	22
6	Deskripsi pemetaan parameter <i>chernoff faces</i> hasil penggerombolan	23
7	Nilai rata-rata peubah numerik hasil penggerombolan optimal ($k = 5$)	24
8	Nilai rata-rata peubah numerik hasil penggerombolan ($k = 4$)	25
9	Nilai rata-rata peubah numerik hasil penggerombolan ($k = 6$)	26

DAFTAR GAMBAR

1	Contoh penerapan <i>chernoff faces</i> pada paket R <i>aplpack</i>	8
2	Persentase amatan yang tersedia dan hilang pada setiap peubah	11
3	Persentase baris data yang dihapus dari <i>dataset</i>	11
4	Histogram penghasilan per bulan ayah	14
5	Histogram penghasilan per bulan ibu	14
6	Histogram uang kuliah tunggal (UKT)	15
7	Histogram jumlah anak dalam keluarga	15
8	Matriks korelasi Pearson antarpeubah	16
9	Distribusi mahasiswa berdasarkan jalur masuk	16
10	Distribusi mahasiswa berdasarkan jenis kelamin	17
11	Distribusi mahasiswa berdasarkan jurusan	17
12	Distribusi mahasiswa berdasarkan fakultas diterima	17
13	Sebaran mahasiswa berdasarkan pulau domisili	18
14	Distribusi mahasiswa berdasarkan pendidikan terakhir ayah	18
15	Distribusi mahasiswa berdasarkan pendidikan terakhir ibu	19
16	Distribusi mahasiswa berdasarkan pekerjaan ayah	19
17	Distribusi mahasiswa berdasarkan pekerjaan ibu	19
18	Distribusi mahasiswa berdasarkan penanggung biaya pendidikan	20
19	Distribusi mahasiswa berdasarkan daya listrik	20
20	Rasio S_W dan S_B untuk menentukan gerombol optimal	21
21	<i>Chernoff faces</i> hasil penggerombolan optimal ($k = 5$)	24
22	<i>Chernoff faces</i> hasil penggerombolan ($k = 4$)	25
23	<i>Chernoff faces</i> hasil penggerombolan ($k = 6$)	26
24	Proporsi jalur masuk berdasarkan hasil penggerombolan	28
25	Proporsi jenis kelamin berdasarkan hasil penggerombolan	28
26	Proporsi jurusan berdasarkan hasil penggerombolan	29
27	Proporsi fakultas diterima berdasarkan hasil penggerombolan	29
28	Proporsi pulau domisili berdasarkan hasil penggerombolan	30
29	Proporsi pendidikan terakhir ayah berdasarkan hasil penggerombolan	30
30	Proporsi pendidikan terakhir ibu berdasarkan hasil penggerombolan	31
31	Proporsi pekerjaan ayah berdasarkan hasil penggerombolan	31

32	Proporsi pekerjaan ibu berdasarkan hasil penggerombolan	32
33	Proporsi penanggung biaya pendidikan berdasarkan hasil	33
34	Proporsi daya listrik berdasarkan hasil penggerombolan	33

DAFTAR LAMPIRAN

1	Kategori awal dan akhir dari peubah-peubah kategorik	43
2	Frekuensi kategori awal dari peubah-peubah kategorik	48
3	Statistik deskriptif peubah-peubah numerik hasil penggerombolan ($k = 5$)	54
4	Distribusi frekuensi dan persentase kategori dari peubah-peubah kategorik berdasarkan hasil penggerombolan ($k = 5$)	55
5	Rangkuman karakteristik masing-masing gerombol	60